

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 07:33:10 ; Search time 4085.76 Seconds  
(without alignments)  
2243.358 Million cell updates/sec

Title: US-09-749-185-1  
Perfect score: 438  
Sequence: 1 atgcgcgagtcggttcaagc.....aggagcagaacccggctga 438

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hlg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	438	100.0	438	6	AX007216	AX007216	438 bp	DNA	linear	PAT 06-SEP-2000
2	438	100.0	1513	1	D50051	Sequence: 1 from Patent WO0000613.				
3	407	92.9	407	6	AX007218	AX007216				
4	327.6	74.8	438	1	AF195772	AX007216.1	GI:9995082			
5	298.2	68.1	407	6	AX007224	Streptomyces griseus				
6	281.2	64.2	15441	1	SCQ11	Streptomyces griseus				
7	265	60.5	566	1	AF195771	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
8	260.8	59.5	408	1	AF195770	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
9	259.8	59.3	407	6	AX007222	1 (bases 1 to 438)				
10	239	54.6	407	6	AX007220	Kraal,B., Luiten,R.G. and Van Wezel,G.P.				
11	113.4	25.9	35284	1	SCE19A	Reducing branching and enhancing fragmentation in culturing				
12	107.4	24.5	38640	1	SCL2	Patent: WO 0000613-A 1 06-JAN-2000;				
13	90.8	20.7	1410	1	STMLACBG	UNIV LEIDEN (NL); KRAAL BAREND (NL);				
14	85.2	19.5	3704	1	SC8A11	(NL); NL ORGANISATIE VOOR WETENSCHAP (NL);				
15	80.8	18.4	37245	1	SC5F2A	LUITEN RUDOLF GIJSBERTUS MARIA				
16	77.2	17.6	45624	6	AX089419	(NL); WEZEL GILLES PHILIPPUS				
17	77.2	17.6	50000	6	AX089416					
18	63	14.4	36583	1	SC5H1					
19	58.6	13.4	438	6	AX007216					
20	58.6	13.4	1513	1	D50051					
21	58.4	13.3	40356	1	SCL6					
22	58.2	13.3	110000	2	LMFLCHR32_06					
23	57.6	13.2	110000	2	LMFLCHR36_07					
24	55.4	12.6	939	1	D87847					
25	55.4	12.6	37200	1	SCE63					
26	55.2	12.6	36849	1	SC1C3					
27	55.2	12.6	65140	6	AX211705					
28	55.2	12.6	123580	1	AF263912					
29	55.2	12.6	125401	6	AX211739					
30	54.8	12.5	80161	1	AY007564					
31	54.8	12.5	80161	6	AL65018					
32	54.8	12.5	215050	1	AL646084					
33	54.6	12.5	10344	1	AE004767					
34	54.4	12.4	1549	1	XANHEMA					
35	54.4	12.4	1549	6	E08655					
36	54.2	12.4	110000	2	LMFLCHR36_31					
37	54	12.3	103843	2	AP004006					
38	53.8	12.3	1778	1	D87846					
39	53.8	12.3	7235	1	SGNUSG					
40	53.8	12.3	122715	2	AC092558					
41	53.6	12.2	407	6	AX007218					
42	53.4	12.2	1972	3	LMA243460					
43	53.4	12.2	1981	6	AR103043					
44	53.4	12.2	1981	6	AR103044					
45	53.4	12.2	1981	6	AR103045					

ALIGNMENTS

RESULT 1

AX007216	AX007216	438 bp	DNA	linear	PAT 06-SEP-2000
LOCUS	Sequence: 1 from Patent WO0000613.				
DEFINITION	AX007216				
ACCESSION	AX007216.1	GI:9995082			
VERSION	Streptomyces griseus				
KEYWORDS	Streptomyces griseus				
SOURCE	Streptomyces griseus				
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
REFERENCE	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
AUTHORS	1 (bases 1 to 438)				
TITLE	Kraal,B., Luiten,R.G. and Van Wezel,G.P.				
JOURNAL	Reducing branching and enhancing fragmentation in culturing				
FEATURES	Patent: WO 0000613-A 1 06-JAN-2000;				
source	UNIV LEIDEN (NL); KRAAL BAREND (NL);				
	(NL); NL ORGANISATIE VOOR WETENSCHAP (NL);				
	LUITEN RUDOLF GIJSBERTUS MARIA				
	(NL); WEZEL GILLES PHILIPPUS				
	VAN (NL)				
	Location/Qualifiers				
	1. 438				
	/organism="Streptomyces griseus"				
	/db_xref="taxon:1911"				

```

gene          1. .408
              /gene="ssgA"
CDS           1. .>408
              /note="ssgA"
              /note="strain ATCC of Streptomyces griseus"
              /codon_start=1
              /transl_table=11
              /product="ssgA"
              /protein_id="CAC07384.1"
              /db_xref="GI:995083"
              /translation="MRESVQAFVMSFLVSELSFRIPVELRYEVDPAIRMTFHLPGDAPVTFWAF
BASE COUNT   68 a 145 c 153 g 72 t
ORIGIN
Query Match   100.0%; Score 438; DB 6; Length 438;
Best Local Similarity 100.0%; Pred. No. 2.3e-47;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcgcgagtcggttcaagcagaggtcatgatgagcttctcgtctccgagagctctcg 60
Db 1 ATCGCGAGTCGGTTCACGACAGAGGTCATGATGAGCTTCTCTCTCCGAGGAGCTCTCG 60

QY 61 ttccgtattccggtgagctccgatacagaggtcgcgatccggtatgccatccggatgacg 120
Db 61 TTCCGTATTCCGGTGGAGCTCCGATACGAGGTCGGCGATCCGATGCCATCCGGATGACG 120

QY 121 ttccacttcccggcgatgccccctgtgacctggcggttccggcgagctgctgctggac 180
Db 121 TTCCACTTCCCGCGATGCCCTGTGACCTGGGCGCTTCGGCCGCGAGCTGCTGCTGGAC 180

QY 181 ggactcaacgcccagcgacgacgagctgtcacatcgcccgccgacgcccggggc 240
Db 181 GGACTCAACGCCCAGCGACGCGACGGCGATGTCACATCGCCGCCGACCCGAGCCGAGGGC 240

QY 241 ctggagatgtccacatcccggtccagggtcgcgcgagacctgctgttccggggcg 300
Db 241 CTGGAGATGTCCACATCCCGGCTCCAGGTCCGGCGGACCGTGCCTGTTCGGCGCGGG 300

QY 301 acggcacccgtgtggttctcctcagccgagacagacagctgctcgcgctcgccaggag 360
Db 301 ACGGCACCCGTGTGGCTTCTTCGACCCGAGCAGACAGCTGCTGCGGCTCGGCCAGGAG 360

QY 361 cacacgtggtgacttcgacgcaacctgagagcactggcgacgactcctcgcgcgag 420
Db 361 CACACGTGGTGACTTCGACGCAACCTGGAGACGCACTGGCGCGCATCTCTCGCCGAG 420

QY 421 gagcagaacccggctga 438
Db 421 GAGCAGAACCGCGCTGA 438

RESULT 2
D50051
LOCUS       D50051 1513 bp DNA linear BCT 19-MAY-1999
DEFINITION Streptomyces griseus DNA for ssgA, complete cds.
ACCESSION  D50051
VERSION    D50051.1 GI:1772323
KEYWORDS   ssgA.
SOURCE     Streptomyces griseus (strain:B2682) DNA.
ORGANISM   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomyces.

REFERENCE  1 (sites)
AUTHORS   Kawamoto,S. and Ensign,J.C.
TITLE     Cloning and characterization of a gene involved in regulation of
            sporulation and cell division of Streptomyces griseus
JOURNAL   Actinomycetol. 9, 136-151 (1995)
REFERENCE  2 (sites)
AUTHORS   Kawamoto,S. and Ensign,J.C.
TITLE     Isolation of mutants of Streptomyces griseus that sporulate in

```

```

nutrient rich media
Actinomycetol. 9, 124-135 (1995)
3 (sites)
Kawamoto,S., Watanabe,H., Hesketh,A., Ensign,J.C. and Ochi,K.
Expression analysis of the ssgA gene product, associated with
sporulation and cell division in Streptomyces griseus
Microbiology (Reading, England) 143 (Pt 4), 1077-1086 (1997)
97286526
9141673
4 (bases 1 to 1513)
Shinichi,K. and Ensign,J.
Cloning and characterization of a gene involved in sporulation and
cell division of Streptomyces griseus
Unpublished (1995)
5 (bases 1 to 1513)
Shinichi,K.
Direct Submission
Submitted' (06-APR-1995) Kawamoto Shinichi, National Food Research
Institute, Biomolecular Transfunction Laboratory; Kannondai 2-1-2,
Tsukuba, Ibaraki 305, Japan (Tel:0298-38-8124, Fax:0298-38-7996)
Location/Qualifiers
1. .1513
/organism="Streptomyces griseus"
/strain="B2682"
/db_xref="taxon:1911"
385. .389
392. .802
/citation=[4]
/codon_start=1
/transl_table=11
/product="ssgA"
/protein_id="BAA21558.1"
/db_xref="GI:2281004"
/translation="MMSFLVSELSFRIPVELRYEVDPAIRMTFHLPGDAPVTFWAF
DKLVLPGQSHLGLDFDGNLEDALGRILAEQNAG"
839. .875
stem_loop 839. .875
stem_loop 863. .968
BASE COUNT 251 a 539 c 493 g 230 t
ORIGIN
Query Match   100.0%; Score 438; DB 1; Length 1513;
Best Local Similarity 100.0%; Pred. No. 1.7e-47;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcgcgagtcggttcaagcagaggtcatgatgagcttctcgtctccgagagctctcg 60
Db 365 ATCGCGAGTCGGTTCACGACAGAGGTCATGATGAGCTTCTCTCTCCGAGGAGCTCTCG 424

QY 61 ttccgtattccggtgagctccgatacagaggtcgcgatccggtatgccatccggatgacg 120
Db 425 TTCCGTATTCCGGTGGAGCTCCGATACGAGGTCGGCGATCCGATGCCATCCGGATGACG 484

QY 121 ttccacttcccggcgatgccccctgtgacctggcgcgagacctgctgctcgcgctcgac 180
Db 485 TTCCACTTCCCGCGATGCCCTGTGACCTGGCGCTTCGGCCGCGAGCTGCTGCTGGAC 544

QY 181 ggactcaacgcccagcgacgacgagctgtcacatcgcccgccgacgcccaggagc 240
Db 545 GGACTCAACGCCCAGCGACGCGCGGCGATGTCACATCGCCGCCGACCCGAGCCGAGGC 604

QY 241 ctggagatgtccacatcccggtccagggtcgcgcgagacctgctgttccggggcg 300
Db 605 CTGGAGATGTCCACATCCCGGCTCCAGGTCCGGCGGACCGTGGCGCTGTTCGGCGCGGG 664

QY 301 acggcacccgtgtggttctcctcagccgagacagacagctcgtgcccgtcgccaggag 360
Db 665 ACGGCACCCGTGTGGCTTCTTCGACCCGAGCAGACAGCTGCTGCGGCTCGGCCAGGAG 724

QY 361 cacacgtggtgacttcgacgcaacctgagagcactggcgacgactggggccactcctcgccgag 420
Db 725 CACACGTGGTGACTTCGACGCAACCTGGAGACGCACTGGGCCGCGCATCTCTCGCCGAG 784

```



```
Qy 421 gagcagaacgcgcgctga 438
|||||
Db 785 GAGCAGACGCCGCTGA 802

RESULT 3
AX007218
LOCUS AX007218 407 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 3 from Patent WO00000613.
ACCESSION AX007218
VERSION AX007218.1 GI:9995084
KEYWORDS
SOURCE Streptomyces griseus.
ORGANISM Streptomyces griseus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 407)
Kraal,B., Luiten,R.G. and Van Wezel,G.P.
Reducing branching and enhancing fragmentation in culturing
filamentous microorganisms
Patent: WO 0000613-A 3 06-JAN-2000;
UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARTA
(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS
VAN (NL)
Location/Qualifiers
source 1..407
/organism="Streptomyces griseus"
/db_xref="taxon:1911"
gene 1..405
/gene="ssgA"
1..>405
/gene="ssgA"
/strain="ATTC of Streptomyces griseus"
/codon_start=1
/transl_table=11
/product="SsgA"
/protein_id="CAC07385.1"
/db_xref="GI:9995085"
/transl_table="MSFLVSEELSFRIPELVYEVGDPYAIRMTFHLPGDAPVTWAFG
RELLDGLNSPSGDVGHIGTPEPEGLGDVHIRLOVGADRALFRAGTAPLVAFLDRFD
KLVPFGHEHTLGFDFGNLELDALRIAEQNAQ"
BASE COUNT 60 a 139 c 142 g 66 t
ORIGIN

Query Match 92.98; Score 407; DB 6; Length 407;
Best Local Similarity 100.0%; Pred. No. 2.2e-43;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 atgagcttcctcgtccgagagctcgttcgtattccggtgagctccgatacag 90
|||||
Db 1 ATGAGCTTCCTCGTCTCCGAGAGCTCTCGTATCCGTATCCGGTGGAGCTCCGATACGAG 60

Qy 91 gtcggcgatcgatgccatccgagtgacgttccacacctcccgcgatgccctctgtgacc 150
|||||
Db 61 GTCGCGATCGGTATGCCATCCGATGACGTTCCACCTTCCCGCGATGCCCTGTGACC 120

Qy 151 tggcgcttcggcgagactcgtgacgggtcacaacgcccgagcgagcgagat 210
|||||
Db 121 TGGCGGTTCCGCGCGAGCTGCTGAGACGGGCTCAACACCGCGAGCGGCGCAT 180

Qy 211 gtgcaatcgcccgagccgagggcctcgagatgctccacatccgactccagctc 270
|||||
Db 181 GTGCACATCGGCCCGACCGCCGAGGGCTCGAGATGTCACATCCGGCTCCAGGTC 240

Qy 271 ggcggagaccgtgcgctgttcgggggagcgagccacgcgtggtggcttccctgacgg 330
|||||
Db 241 GCGCGGACCGTGGCTGTTCCGGCGGGGAGCGGACCGCTGGTGGCTTCTTCGACCGG 300

Qy 331 acggacaagctcgtgcgcctcggccagagacacgcgtgggtgacttcgagcggaacctg 390
|||||
Db 301 ACGGACAAGCTCGTGCCTCGGCGAGGACACACGCTGGGTGACTTCGACGGGAACCTG 360
```

```
Qy 391 gagcagcactggcgcccatctcgcgagagcagagcagcgcgcgctg 437
|||||
Db 361 GAGGACGCACTGGCGCGCATCTCGCCGAGGAGCAGACGCCGCGCTG 407

RESULT 4
AF195772
LOCUS AF195772 438 bp DNA linear BCT 01-NOV-2000
DEFINITION Streptomyces netropsis SsgA (ssgA) gene, complete cds.
ACCESSION AF195772
VERSION AF195772.1 GI:11066162
KEYWORDS
SOURCE Streptomyces netropsis.
ORGANISM Streptomyces netropsis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 438)
van Wezel,G.P., Rousseau,C. and Kraal,B.
Cloning and sequencing of the Streptomyces netropsis ssgA gene
Unpublished
REFERENCE 2 (bases 1 to 438)
AUTHORS van Wezel,G.P.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Biochemistry, University of Leiden, PO Box
9502, Leiden 2300 RA, Netherlands
Location/Qualifiers
source 1..438
/organism="Streptomyces netropsis"
/db_xref="taxon:55404"
gene 1..438
/gene="ssgA"
1..438
/gene="ssgA"
CDS 1..438
/strain="ATTC of Streptomyces griseus SsgA; possible cell
division protein"
/codon_start=1
/transl_table=11
/product="SsgA"
/protein_id="NAG28483.1"
/db_xref="GI:11066163"
/translation="MRESVQAEVLSFLVSEELSFRIPELVYETRPDYAVRMTFHLPG
DAPVTWAFWAGRELLDGLNRPISGDVGHIGTPEPEGLGDVHIRLOVGADRALFRAGAP
PLVAFLDRTDKSVPLGQETGLGDFEDSLAEALGKILAEQNAQ"
BASE COUNT 70 a 158 c 142 g 68 t
ORIGIN

Query Match 74.8%; Score 327.6; DB 1; Length 438;
Best Local Similarity 84.2%; Pred. No. 2.9e-33;
Matches 369; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 1 atgcgcgagtcggttcgaagcagagtgatgatgagcttcctcgtccgagagctctcg 60
|||||
Db 1 ATGCGGAGTCGGTTCAAGCAGAGGTCCTGATGAGCTTCTCGTCTCCGAGGAGCTCC 60

Qy 61 ttccgtattccggtgagctccgatacaggtcggcgatccgatgccatccgagatgacg 120
|||||
Db 61 TTCAAGATCCCGATCGAACTGGATACGAGACCCGGGATCCCTACCGGTGCGGATGACC 120

Qy 121 ttccaccttcccgcgatgccctgtgacctggcggttcgcgcgagctgctgctggac 180
|||||
Db 121 TTCCACCTCCCGGAGACGCGCTGTGACCTGGGCGTTCCGCGCGGAGCTGCTGCTGAC 180

Qy 181 gggctcaacagcccgagcggcgacggcgatgtgcacatcgccccgacccgagccccgagggc 240
|||||
Db 181 GGGATCAACCGCCCGAGCGGCGACGGCGACGCTCCACATCGCCCCCGACGCCCGAGGGC 240

Qy 241 ctgcgagatgtccacatccggctccaggtcggcgagccctgctgttccggcgggg 300
|||||
Db 241 CTGCGGAGCTCTCCATCCGCTCCAGGTGGCGCGGACCGCGCCCTCTTCGTCGAGGC 300

Qy 301 acggcaccgcgtgtggttcctcctcgcgagcggacgagctgctgcgctcgccgagga 360
|||||
```

```

Db 301 GCGCCGCGCTGGTCTCTTCGACCGACGAGAAAGTCGGTGGCTCGGTCTCGACGCG 360
Qy 361 cacactcgggtgacttcacgcgaacctggaggacgcactggcgccgctcctccgcgag 420
Db 361 CAGACTCTGGGTGACTTCGAGGACAGCTGGAGCGCGCTCGGCAAGATCTCTCGCCGAG 420
Qy 421 gagcagaacgccgctga 438
Db 421 GAGCAGACGCGGCTGA 438

RESULT 5
LOCUS AX007224 407 bp DNA linear PAT 06-SRP-2000
DEFINITION Sequence 9 from Patent WO0000613.
ACCESSION AX007224
VERSION AX007224.1 GI:9995090
KEYWORDS Streptomyces netropsis.
SOURCE Streptomyces netropsis.
ORGANISM Streptomyces netropsis.
REFERENCE Kraal,B., Luiten,R.G. and Van Wezel,G.P.
AUTHORS Reducing branching and enhancing fragmentation in culturing
TITLE filamentous microorganisms
JOURNAL Patent: WO 0000613-A 9 06-JAN-2000;
UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GLJSBERTUS MARIA
(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS
VAN (NL)
FEATURES
source Location/Qualifiers
1..407
/organism="Streptomyces netropsis"
/db_xref="taxon:55404"
1..405
/gene="ssga"
1..>405
/gene="ssga"
/codon_start=1
/transl_table=11
/product="Ssga"
/protein_id="CAC07388.1"
/db_xref="GI:9995091"
/translation="MSFLVSEELSPKIPVELRYETRPVAVRMTFLPGDAPVTWAFG
RELLDGINRPSGCDGVHIAPTDPEGLSDVSIQLVGADRALFRAGAPLVAFLDRDT
KSPVIGQETLGFDFEDSLKALGKIILAEQNAG"
BASE COUNT 63 a 151 c 131 g 62 t
ORIGIN

Query Match 68.1%; Score 298.2; DB 6; Length 407;
Best Local Similarity 83.3%; Pred. No. 1.7e-29;
Matches 339; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 31 atgagctctcgtctccgaggagctctcgttcggtattccggtggagctcccgatacag 90
Db 1 ATGAGCTCTCTCTCCGAGGAGCTCTCTCAAGATCCAGTCGAGACTCGGATACGAG 60
Qy 91 gtcggcgatccgtatgccatcccgatgacgttcacacttcccggcgatgccctgtgacc 150
Db 61 ACCCGGATCCCTACCGGGTGGGATGACCTTCACCTCCCGGAGACGGCGCTGTGACC 120
Qy 151 tggcggttcggcgagctgctgtgacggctcaacagccgagcgagcgagcgat 210
Db 121 TGGCGGTTCGGCGGGAGCTGCTCTCGACGGGATCAACCGCGGAGCGGCGGCGAC 180
Qy 211 gtgcacatcgcccgacgcagcgaggcgctcgagatgtccacatccggctccaggtc 270
Db 181 GTCCACATCGCCCGACCGACCGCCGAGGGCTGTCTCGACGCTCTCCATCCGGCTCCAGGTG 240
Qy 271 ggcgcgacogtgcgtgtttcccgggcgagcgacccgctggtggttccctcgaccgg 330
```

```

Db 241 GGCGCGGACCGCGCCCTCTTCGTCGAGCGCCCGCCGCTGGTCTCTCGACCGC 300
Qy 331 acggacaagtctcgtcgcgtcgcgcaggagcacagcctgggtgacttcgacggcaacctg 390
Db 301 ACGGACAAGTCGGTGGCGCTCGGTAGGAACAGACTCTGGTGACTTCGAGGACAGCCTG 360
Qy 391 gaggacgcactggcgccgcatctcgcgcggaggagcagaacgcggctg 437
Db 361 GAGGCGCGCTCGGCAAGATCTCTCGCGGAGGAGCAGAGCGCGGCTG 407

RESULT 6
LOCUS SCQ11 15441 bp DNA linear BCT 08-JUL-1999
DEFINITION Streptomyces coelicolor cosmid Q11.
ACCESSION AL096823
VERSION AL096823.1 GI:5457267
KEYWORDS abab; bldA regulation; cysA; cystathionine/methionine
gamma-synthase/lyase; gntR-family; integrated element; korSA;
lysR-family; membrane protein; phosphotyrosine protein phosphatase;
pra; pSAM2; ptpA; ssgA; thiamine biosynthesis; thic;
transcriptional regulator; trsA.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
REFERENCE Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
AUTHORS Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 15441)
Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
JOURNAL Unpublished
MEDLINE 2 (bases 1 to 15441)
REFERENCE Seeger,K. and Harris,D.
AUTHORS Unpublished
JOURNAL 3 (bases 1 to 15441)
REFERENCE James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
AUTHORS Direct Submission
TITLE Submitted (08-JUL-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S.coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.go.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
```

sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid Q11 Cosmid Q11 lies between H24 and D78 on the AseI-Q genomic restriction fragment.

## FEATURES

```

source
    location/Qualifiers
    1..15441
    /organism="Streptomyces coelicolor A3(2)"
    /strain="A3(2)"
    /db_xref="taxon:100226"
    /clone="cosmid Q11"
    /complement(1..289)
    /gene="SCQ11.01c"
    /complement(1..289)
    /gene="SCQ11.01c"
    /note="SCQ11.01c, hypothetical protein, partial CDS, len:
    >95 aa; unknown function, previously sequenced as
    TR:Q53870 (EMBL:U37580), S.coelicolor hypothetical protein
    ORF5 downstream of phosphotyrosine protein phosphatase
    (159 aa). Weak similarity to the N-terminus of TR:Q69888
    (EMBL:AL023797) Streptomyces coelicolor hypothetical
    protein (172 aa), fasta scores: opt: 158 z-score: 206.4
    E(): 0.00036, 46.2% identity in 65 aa overlap. Continues
    as SCH24.40c in cosmid H24 (EMBL:AL049826)"
    /codon_start=1
    /transl_table=11
    /label="SCQ11.01c"
    /product="hypothetical protein"
    /protein_id="CAB46956.1"
    /db_xref="GI:5457268"
    /translation="MTVRPVKRTARAVLDGDHLILIKRTKGVDPYVWTPGGVPE
    DDTVTVDALHREYIEELGAKISDVPCFVDTVEHIGDGGATGKVKQHFVC"
    /complement(1..4279)
    /note="sequence corresponding to EMBL:U37580 from 1 to
    4279"

misc_feature
    1..98
    /note="overlap with Streptomyces coelicolor cosmid H24
    (EMBL:AL049826) from 41528 to 41625"
    /complement(360..1265)
    /gene="SCQ11.02c"
    /note="abaB"
    /complement(360..1265)
    /gene="SCQ11.02c"
    /note="SCQ11.02c, abaB, probable lysR-family
    transcriptional regulator, len: 301 aa; previously
    sequenced as TR:Q53869 (EMBL:U37580) S.coelicolor
    lysR-like protein (301 aa). Highly similar to
    SW:ARAB_STRAT (EMBL:X91393), arab, Streptomyces
    antibioticus transcriptional regulatory protein (301 aa),
    fasta scores: opt: 1581 z-score: 1813.2 E(): 0, 80.3%
    identity in 300 aa overlap. Similar to many other
    transcriptional regulators. Contains probable
    helix-turn-helix motif at aa 17-38 (score 1659, +4.84 SD).
    Contains Pfam match to PF00126 HTH_1, Bacterial regulatory
    helix-turn-helix protein, lysR family and PS00044
    Bacterial regulatory proteins, lysR family signature"
    /codon_start=1
    /transl_table=11
    /label="abaB"
    /product="putative lysR-family transcriptional regulator"
    /protein_id="CAB46957.1"
    /db_xref="GI:5457269"
    /translation="MDLALLRTEVTVHRAGSFTRAAALLGLSOPAVTSQIRTLEROIG
    RPLFLRQAGVPTTIGDELAKHAPHLDAVLAEIATGLEDSDTLKTLHLAGPPTTA
    ERALPALGELTGDDGQAFALRSGFNAETLEGAGHHDLAGITRPRGALHTATPL
    CDBVHLVATPHWAERAGVEDVDTDAALKHVPVVEHESLFFVGRYWASVEDARPA
    SPATVVPDLRAVLACAVAGLAVLPFYLCALERGDVVALHDPVPPLRTYFLVY
    RTGTAMPHIARAHWLLRAADWN"
    /complement(831..1256)
    /gene="SCQ11.02c"
    /note="Match to PF00126 HTH_1, Bacterial regulatory
    helix-turn-helix protein, lysR family Score 145.96"
    /complement(1125..1217)

misc_feature
```

```

    /gene="SCQ11.02c"
    /note="PS00044 Bacterial regulatory proteins, lysR family
    signature"
    /complement(1518..2696)
    /gene="SCQ11.03c"
    /note="cysA"
    /complement(1518..2696)
    /gene="SCQ11.03c"
    /note="SCQ11.03c, cysA, probable cystathionine/methionine
    gamma-synthase/lyase, len: 392 aa; previously sequenced as
    SW:CYS_A_STRCO (EMBL:U37580), cysA, S.coelicolor putative
    cystathionine gamma-lyase (392 aa). Similar to many e.g.
    SW:MEGL_PSEPU (EMBL:D88554), mdeA, Pseudomonas putida
    methionine gamma-lyase (398 aa), fasta scores: opt: 608
    z-score: 653.6 E(): 4.5e-29, 36.9% identity in 404 aa
    overlap. Highly similar to TR:Q53668 (EMBL:X91393)
    Streptomyces antibioticus hypothetical protein found
    upstream of the abaB gene (232 aa) (87.1% identity in 232
    aa overlap). Contains Pfam match to PF01053
    Cys Met Meta_PP, Cys/Met metabolism PLP-dependent enzyme
    and PS00868 Cys/Met metabolism enzymes pyridoxal-phosphate
    attachment site"
    /codon_start=1
    /transl_table=11
    /label="cysA"
    /product="putative cystathionine/methionine
    gamma-synthase/lyase"
    /protein_id="CAB46958.1"
    /db_xref="GI:5457270"
    /translation="MSDATDTSAGTGGERSASAPGDTGTRAVRAGLPEPVKHEPTLPG
    PVFAHFHLPDGTPTGYGRDNPTWRLSAIGELEAFGEAGVETLVFASGMAAIS
    SVLFSQLRAGDTAVLPDGYOALPLVRAOLEAYGIEVTRAPTGRDOLDVLDGAKLLW
    IETSPNGLDCDVRRLVEAAHAGGALVAVDNTLATPLGORPLEGADFVSASCTKOL
    TGHGVLLCYVAGRDAGACAAAVRRWRKIYCAIGCPMEAWLAHRSIATLQLRYDRQDST
    ALKVAALRTPEITGLRYPGLDDPSHKRGMQMLRYGCVSFTLPSRAARFLDA
    LRLVEGATSGFGRVSTAEARRGRWGDAVPEGFIRLSVGAEDPDDLVALDLRALDETTE
    "
    /complement(1533..2624)
    /gene="SCQ11.03c"
    /note="Match to PF01053 Cys_Met_Meta_PP, Cys/Met
    metabolism PLP-dependent enzyme Score 268.52"
    /complement(2031..2075)
    /gene="SCQ11.03c"
    /note="PS00868 Cys/Met metabolism enzymes
    pyridoxal-phosphate attachment site"
    /complement(2693..3187)
    /gene="SCQ11.04c"
    /note="ptpA"
    /complement(2693..3187)
    /gene="SCQ11.04c"
    /note="SCQ11.04c, ptpA, low molecular weight
    protein-tyrosine-phosphatase, len: 164 aa; previously
    sequenced as SW:PTPA_STRCO (EMBL:U37580), ptpA,
    S.coelicolor low molecular weight
    protein-tyrosine-phosphatase (164 aa). Contains Pfam match
    to PF01451 LMWPC. Low molecular weight phosphotyrosine
    protein phosphatase"
    /codon_start=1
    /transl_table=11
    /label="ptpA"
    /product="low molecular weight
    protein-tyrosine-phosphatase"
    /protein_id="CAB46959.1"
    /db_xref="GI:5457271"
    /translation="MTYRVCFVCTGNTICRSPMAEAVFRARVEDAGLHLYEASAGTG
    GWHEGADPTAEAVLHDGHLDAHQVQQSWFSRLDLVVALDAGHLRLRLAPT
    ERDAAKVRLRLSDPAVAGDLDPDPYGYGRDGFECLEWVEAASGTLAAVREQVE
    GRAA"

Query Match          54.2%; Score 281.2; DB 1; Length 15441;
Best Local Similarity 77.6%; Pred. No. 9.9e-28;
Matches 340; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
```



```

/transl_table=11
/product="SsgA"
/protein_id="CAC07387.1"
/db_xref="GI:9995083"
/translation="MSFLVSEELSFRIPELVRLTCDPYAVRLTFHLPDGPVYTWAFG
RELLIDGGPPCGDGDVHIAPDPETFGVEVLIRLQVGSQAMFRVGTAPLVAFLDRTD
KIVPLGGRSLADFALLDRLALAEQNAG"
56 a 135 c 145 g 71 t
BASE COUNT

```

Query Match	59.3%; Score 259.8; DB 6; Length 407;
Best Local Similarity	77.4%; Pred. No. 1.3e-24;
Matches 315; Conservative	0; Mismatches 92; Indels 0; Gaps 0;
Qy	31 atgagctctcgttcgcgaggagtctcgttccqattccggtggagctccgatacga 90       
Db	1 ATGAGCTTCCTCGTTCGGAGAACTCCTTCCTCGATTCCGGTGAGACTCGCTTACGAG 60       
Qy	91 gtccggcatccgatgccatccgatgcgttccacettcccggcgatgcccttgtacc 150       
Db	61 .ACCTGTATCCTACGCCGTGGGTGACCTTTTCATCTGCCCGAGATGCCCGGTGACC 120       
Qy	151 tggcggttcgcgcgagactcgtcgacgggtcaacagccgagcggcgacggcgat 210       
Db	121 TGGCGGTTCCGGCGGGAGTTCTCATCACGGAGGTCGCGCGCCGTGGGGAGCGGGAC 180       
Qy	211 gtgcacatcgcccgacgcgagccgcgagggctcgagatgtccacatccggtccaggtc 270       
Db	181 GTCCACATCGCGCCGCCGACCCGGAGACGTTCCGGAGGTTCTGATCCGCCTGCAGGTG 240       
Qy	271 ggcgcgacctgcgctgtttccggcgccgagcacgcagctggttggttctctcgacgg 330       
Db	241 GGGAGCAGCAAGGCGATGTTCCGGGTTCGGCACGCGCGCGCTGGTGGCTTCTCTGGACGC 300       
Qy	331 acggacaagctcgtgcccgtcggccagagcacgcgtgggtgacttcgacggcaaacctg 390       
Db	301 ACGGACAAGATCGTCCCGCTGGGGCAGGAGGTTCCCTTCGCCGACTTCGAGCGCCCTGCTC 360       
Qy	391 gaggacgactggccgcacatcctgcgcgagggagcagaacgcgcgctg 437       
Db	361 GACGAGGCGCTGGACCGCATCTGTCGCCAGGAGCAGAACGCCGCGTG 407       

```

RESULT 10
AX007220
LOCUS      AX007220          407 bp      DNA      linear      PAT 06-SEP-2000
DEFINITION Sequence 5 from Patent WO0000613.
ACCESSION  AX007220
VERSION     AX007220.1  GI:9995086
KEYWORDS
SOURCE      Streptomyces albus G.
ORGANISM    Streptomyces albus G.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1 (bases 1 to 407)
AUTHORS     Kraal,B., Luiten,R.G. and Van Wezel,G.P.
TITLE        Reducing branching and enhancing fragmentation in culturing
              filamentous microorganisms
              Patent: WO 0000613-A 5 06-JAN-2000;
              UNIV LEIDEN (NL); KRAAL, BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA
              (NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS
              VAN (NL)
JOURNAL
FEATURES    Location/Qualifiers
             1..407
              /organism="Streptomyces albus G"
              /db_xref="taxon:1962"
             1..405
              /gene="ssgA"
             1..>405
              /gene="ssgA"
              /note="strain ATCC of Streptomyces albus G"
              /codon_start=1

```

JOURNAL		Submitted (16-JUL-1999) Streptomyces coelicolor sequencing project Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK	
COMMENT		Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S.coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nh.hgo.jp/ fun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid E19a lies between and overlaps cosmids E6 and E56 on the AseI-E genomic restriction fragment.	
FEATURES		Location/Qualifiers 1..35284 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid E19a" 1..97 /gene="SCE19A.01" 1..97 /gene="SCE19A.01" /note="SCE19A.01" >31 aa; unknown function, similar to members of the alkyl hydroperoxide reductase C/thiol-specific antioxidant family e.g. TR:O53226 (EMBL:AL021185), bcp, Mycobacterium tuberculosis bacterioferritin comigratory protein (157 aa), fasta scores; opt: 124 z-score: 238.1 E(): 6.2e-06, 63.3% identity in 30 aa overlap" /codon_start=2 /transl_table=11 /label="SCE19A.01" /product="hypothetical protein" /protein_id="CAB50982.1" /db_xref="GI:5531350" /translation="IVVDEQGVKVERLYNVRATGHVAKIKDLGI" 140..226 /note="rRNA Leu anticodon TAG, Cove score 69.78" /product="rRNA-Leu" complement(243..845) /gene="SCE19A.02c" complement(243..845) /gene="SCE19A.02c" /note="SCE19A.02c, conserved hypothetical protein, len: 200 aa; unknown function, similar to many e.g. SW:Y03Q_MYCTU (EMBL:273902) Mycobacterium tuberculosis hypothetical protein (204 aa) fasta scores; opt: 633	
SOURCE		Source 1..35284 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid E19a" 1..97 /gene="SCE19A.01" 1..97 /gene="SCE19A.01" /note="SCE19A.01" >31 aa; unknown function, similar to members of the alkyl hydroperoxide reductase C/thiol-specific antioxidant family e.g. TR:O53226 (EMBL:AL021185), bcp, Mycobacterium tuberculosis bacterioferritin comigratory protein (157 aa), fasta scores; opt: 124 z-score: 238.1 E(): 6.2e-06, 63.3% identity in 30 aa overlap" /codon_start=2 /transl_table=11 /label="SCE19A.01" /product="hypothetical protein" /protein_id="CAB50982.1" /db_xref="GI:5531350" /translation="IVVDEQGVKVERLYNVRATGHVAKIKDLGI" 140..226 /note="rRNA Leu anticodon TAG, Cove score 69.78" /product="rRNA-Leu" complement(243..845) /gene="SCE19A.02c" complement(243..845) /gene="SCE19A.02c" /note="SCE19A.02c, conserved hypothetical protein, len: 200 aa; unknown function, similar to many e.g. SW:Y03Q_MYCTU (EMBL:273902) Mycobacterium tuberculosis hypothetical protein (204 aa) fasta scores; opt: 633	
BASE COUNT		57 a 150 c 139 g 61 t	
ORIGIN		/transl_table=11 /product="SsgA" /protein_id="CAC07386.1" /db_xref="GI:995087" /translation="NSFLVSEELAFRIPVELRYETDYPVAVRLTFHLPLDAPVTWVFG RELVEGLDAGDGVPCVPGQTATREVHITLQVGEALFRVGRAPLAFALDRFD QGLSGSERAHAFDSHDDALNSLAEOASG"	
Query Match		54.6%; Score 239; DB 6; Length 407;	
Best Local Similarity		74.2%; Pred. No. 6e-22;	
Matches 302; Conservative		0; Mismatches 105; Indels 0; Gaps 0;	
Qy	31	atagctctctctccagagagctctcggttcgtattccggtgagctccgatacag	90
Db	1	ATGAGCTTCCTGCTCCGAGGAGCTCGGCTTCCGCATCCCGGTGGAGCTGCGGTACGAG	60
Qy	91	gtcgggacatcgatcgatccgagtagcgttccaccttcccgcgatgccctgtgacc	150
Db	61	ACCGTCGATCCGTACCGGTGCGGTGAGCTTCCACCTCCCGGAGAGCCCGGTACCC	120
Qy	151	tggcgcttgcgcgcgagctgtgctgtggacggttcacacagcccgagcgagcgagat	210
Db	121	TGGGTCTTCGGCGGTGAACCTGTGTCGAGGAGTCTTGGACGCCCGCGGCGACGC	180
Qy	211	gtcacatcgcccgacgagccgagggcctgcgagatgtcacatccggctccaggtc	270
Db	181	GTCCGGTCTGCCCCGTGGCGGACAGCCACACGAGGTGCATCACCCTCCAGGTC	240
Qy	271	ggcgcgacgtcgctgttccggcgggagcgacacgcgtgtggttctctgcacgg	330
Db	241	GGCTCCGAGCAGCGCTCTTCCGCTGCGCAGGCGCGCTGCTGCCCTTCTCGACCGC	300
Qy	331	acggacaagctctgcgcgtcggccagggacacacgcgtgggtgacttcgcagcaactc	390
Db	301	ACCGACAGCGGCTGTGCTGCGCAGCAGCGGCGGACACGCCGACTTCGACAGCCACCTC	360
Qy	391	gagagcactggcgacatctccgagagcagaacgcgcgtg	437
Db	361	GACGACGCTCTGACCCGACGCTCTCCGAGGAGCAGAGCGCGGCTG	407
RESULT 11		SCE19A	
LOCUS		35284 bp DNA linear BCT 16-JUL-1999	
DEFINITION		Streptomyces coelicolor cosmid E19a.	
ACCESSION		AL096852	
VERSION		AL096852.1 GI:5531349	
KEYWORDS		4-hydroxyphenylpyruvate dioxygenase; ABC transporter; amino acid permease; asc-family; cysteine synthase; nicotinamide; osmoregulation; osmotolerance; oxidoreductase; permease; phosphoenolpyruvate-dependent; protease; regulator; ribonuclease P4; rph; sugar phosphotransferase system; transcriptional regulator; transposase; rRNA Leu; zinc metalloproteinase. Streptomyces coelicolor A3(2). Streptomyces coelicolor A3(2).	
SOURCE		Streptomyces coelicolor A3(2).	
ORGANISM		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 35284) Redenbach,M., Kieser,H.M., Denapalte,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome Mol. Microbiol. 21 (1), 77-96 (1996) 97000351 2 (bases 1 to 35284) Seeger,K. and Harris,D. Unpublished JOURNAL 3 (bases 1 to 35284) James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A. Direct Submission TITLE	
JOURNAL		Mol. Microbiol. 21 (1), 77-96 (1996)	
MEDLINE		97000351	
REFERENCE		2 (bases 1 to 35284) Seeger,K. and Harris,D. Unpublished JOURNAL 3 (bases 1 to 35284) James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A. Direct Submission TITLE	















jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid 5F2A.

#### FEATURES

##### source

Location/Qualifiers

1. .37245  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 5F2A"

##### misc\_feature

1. .105  
/note="Nominal overlap with cosmid 4C6"

##### gene

complement(83. .859)  
/gene="SC5F2A.01c"  
/note="ddah"

##### CDS

complement(83. .859)  
/gene="SC5F2A.01c"  
/note="SC5F2A.01c, ddah, dimethylarginine dimethylaminohydrolase, len: 258 aa; hydrolyses asymmetrically methylated arginine with preference for dimethylated arginine over monomethylated arginine (experimental). Similar to TR-DI038106 (EMBL:AB001915) NG,NG-dimethylarginine dimethylaminohydrolase from Homo sapiens (285 aa) fasta scores; opt: 323, z-score: 384.7, E(): 4.1e-14, (35.7% identity in 272 aa overlap)."  
/codon\_start=1  
/transl\_table=11  
/label=ddah

/product="dimethylarginine dimethylaminohydrolase"

/protein\_id="CAB40668.1"

/db\_xref="GI:4584465"

/db\_xref="SWISS-PROT:Q9X7M4"

/translation="WPSKALVRPSPRLAEGLVTHVERKVDHGLALBQWDAYVEAL  
GAGWETLEVDADCPDSVFVEDAVVFNVALTRPGAESRRAETAGVEAVARLG  
CSVNWWEGETLDGVDLKIGDTIVYGRGRTNAGVQQLRAAFPEPLGARVYAVPVSK  
VHLKSAVTPDGTIVGHIPLDTPSPLEPRLPVEESGAHVLLGGSRLLMAASAP  
KTAELLADLIGHEPVLVDIGEFKLEGVCVCLSVRLRELYD"  
complement(976. .3234)  
/gene="SC5F2A.02c"

##### gene

complement(976. .3234)

##### CDS

complement(976. .3234)  
/gene="SC5F2A.02c"  
/note="SC5F2A.02c, UvrA-like ABC transporter, len: 752aa; similar to many egs. TR:086699 (EMBL:AL031515) UvrA-like protein from Streptomyces coelicolor (796 aa) fasta scores; opt: 2047, z-score: 2164.0, E(): 0. (51.9% identity in 776 aa overlap) and SW:UVR\_A.ECOLI UvrA, excinuclease ABC subunit A from Escherichia coli (940 aa) fasta scores; opt: 634, z-score: 669.2, E(): 5.8e-30, (33.9% identity in 938 aa overlap). Contains two Pfam matches to entry PF00005 ABC\_tran, ABC transporter and two Prosite matches to PS00017 ATP/GTP-binding site motif A (P-loop). Also contains possible membrane spanning hydrophobic regions."

/codon\_start=1

/transl\_table=11

/label=SC5F2A.02c

/product="UvrA-like ABC transporter"

/protein\_id="CAB40669.1"

/db\_xref="GI:4584466"

/db\_xref="SPTREMBL:Q9X7M5"

/translation="MSEFISITGARENNDQVTLRIKPLGRLVFTGVSGSKSVVFD  
TAVESRRQLNETFTWFRNLKPYRPHADALEGLTPAIVDQRPVGGHSRSTVGTM  
TDTHSVLRVLFSSRHGTPGAGGATAYSFNDPSGMCPCGDGLGRRVQDRIILDPARSL  
ADGAVRFPFPAAGTQGTQYNTNEELDTCKPVGDETAERAFILMRGPKSVTVSSGG  
GTWSTYEGLADRFERYLYLKRDLSCNSETRDLRVGFLVEARPCDGGARLNAALAS  
RIDGHSIADCSRMQITDLIAVLKRGIDDPVALPVAGAAVAALERVEAIGLYLSLDRET

ATLGGEGORLKTVRHGLSSLTGMTYIFDEPSVGLHPRVGRGLRGLDRLRLRDKGNVTL  
VREHDPVIALADHVVDGPRAGADGGRVFEPTAGLAASDTLTGRCIGRRTAVKDT  
VRAPTEGLWVGAERHNDREVTVAFTGLVLTAVTGVAGSGKSTLVAELTGAHPDAVV  
DQSAIGTSRSTPATYILGIMDTVRKVFARTEGAEPFFSAGACGTCGEGIIHTD  
LAFMDPVTTTCHDCGRRFREVELRTVDGSRVADVLAMTAGALGFFSDPGVRRRLR  
ALRDVGLTYLTGQPLSGRQRIRKLATRLHRTGAVYVLDPTTGLHNSDVBGLL  
ALLRDLDAGNTVVVVENHLDVVAHADRVLDLPGDGRDGRVIEGTTPRELLAARG  
STAHLRRATRR"

complement(1090. .1803)

/gene="SC5F2A.02c"

/note="Pfam match to entry PF00005 ABC\_tran, ABC

transporter, score 101.40, E-value 1.7e-26."

complement(1759. .1782)

/gene="SC5F2A.02c"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)."

complement(2002. .3162)

/gene="SC5F2A.02c"

/note="Pfam match to entry PF00005 ABC\_tran, ABC

transporter, score 35.60, E-value 4e-08."

complement(3118. .3141)

/gene="SC5F2A.02c"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)."

complement(3245. .4879)

/gene="SC5F2A.03c"

complement(3245. .4879)

/gene="SC5F2A.03c"

/note="SC5F2A.03c, probable ABC transporter, len: 544aa; similar to many involved in antibiotic resistance egs. TR:Q54381 (EMBL:X79146) from the lincomycin-production gene cluster of Streptomyces lincolnensis 78-11 (578 aa) fasta scores; opt: 1624, z-score: 1696.3, E(): 0. (50.3% identity in 541 aa overlap). Contains two Pfam matches to

entry PF00005 ABC\_tran, ABC transporter, two Prosite matches to PS00017 ATP/GTP-binding site motif A (P-loop) and Prosite match to PS00211 ABC transporters family signature. Also contains a possible coiled-coil region around 250aa."

/codon\_start=1

/transl\_table=11

/label=SC5F2A.03c

/product="putative ABC transporter"

/protein\_id="CAB40670.1"

/db\_xref="GI:4584467"

/db\_xref="SPTREMBL:Q9X7M6"

/translation="WSTSPSTSLTSFAWPDGTEVFDGLVARGPGRTGLVGLNGA  
GKSTLLKLIAGRLTPADSTVRVAGQVYLPQNTLTLRVDALGIDGRRALHAIE  
AGDVEHFETVGDWDVEERALATIGELGLDHIGLDRTVGVSGESVLLRLAAIL  
RPPVDLLDEPTNLDVYARRLYAAVESWPGVWVSHDRELLRDVQDIADLRAGSV  
TWYGGNLTAYPEALAVEQEAERMYRVAESDLRRKRELADAQVYLARRRYGQKMYD  
TKREPRVWKLKARTAQSGAGKYRIMHEKLAGAKERLDDAVEAVRDDDEIRVDLPYT  
AVPPGRTVLTALALELAVARGAGGLDLHGPERIALLRNGAGKTLTLRTVAGELAPV  
AGETAHVPLREFLPQRLDVLDDGLTVAENVAFPGATNNRIRALLARLFRGARADQ  
QAATLSGGERFRAALAAALMAEPQLMLMDEPTNNLMASVRLTGALAEVAGALLY  
ASHDLPLFLESLGITRWLLLEGELEITPEAVGYS"

complement(3290. .3787)

/gene="SC5F2A.03c"

/note="Pfam match to entry PF00005 ABC\_tran, ABC

transporter, score 122.30, E-value 9e-33"

complement(3743. .3766)

/gene="SC5F2A.03c"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)."

complement(4232. .4783)

/gene="SC5F2A.03c"

/note="Pfam match to entry PF00005 ABC\_tran, ABC

transporter, score 143.80, E-value 3.1e-39"

complement(4403. .4447)

/gene="SC5F2A.03c"

/note="PS00211 ABC transporters family signature."

complement(4739. .4762)

/gene="SC5F2A.03c"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)."

4942. .5850

/gene="SC5F2A.04"

```
CDS          4942. .5850

Query Match      18.4%; Score 80.8; DB 1; Length 37245;
Best Local Similarity 54.5%; Pred. No. 0.031;
Matches 216; Conservative 0; Mismatches 162; Indels 18; Gaps 2;

Qy 25 gtcagtgagcttcctcgttcgcgagagctctcgttcggtattccggtggagctccga`84
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6305 GTCAGGCGGCTCTCGTCGGCGCGCGCGGATGCGGAGCATTCGCCGACCCCTGCAC 6246

Qy 85 tacgaggtcggcgatcgatgcacatcgatgacgtttcaaccttccggc----- 135
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6245 TACGACCGCGCCGACCGCTTCGCGGTCGCGATGACCTTCCGCGCCCGGCCACCCCTCGAG 6186

Qy 136 gatgccctgtgacctggggttcggcgagctgctgtaggggtcctcaacagcccg 195
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6185 GCGGTGAGGTCTGCTGGACCTTCTCCGCGAGCTGTCTATCGCCGGGATGCGAGGCCG 6126

Qy 196 agcggcgacggcgatgtgcacatcgcccgaccgagcccgagggcctcggagatgtccac 255
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6125 AACGCCACGGCGACGTCCGGGTGCGCCGCTACGCGTACGACCGCACCG-----TC 6075

Qy 256 atccggtccaggtcggcgagaccgtgacgtgtccggcggggacggacccgctggtg 315
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6074 CTGGAGTTCCACGCGCGCCGAGGACCGCGGTGATCCACGTCCGCTCGGGCGAGCTGCGC 6015

Qy 316 gcgttcctcgaccggacgacaagctcgtccgctcgccagagacacagcgtgggtgac 375
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6014 CGGTTCTTCAGGCGCGCGGAGCTGTCGCGTGGGTCTCGAGACACCTCCACGCTGGAC 5955

Qy 376 ttgacggcaacctggaggacgactgggcgcac 411
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5954 CTGGACCAACGACCTGGCGGAGCTGATGCGCGGGAGC 5919
```

Search completed: July 18, 2002, 11:21:23  
Job time: 13693 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 08:28:55 ; Search time 646.57 Seconds  
(without alignments)  
1163.074 Million cell updates/sec

Title: US-09-749-185-1  
Perfect score: 438  
Sequence: 1 atgcgcagtcggttaacg.....aggagcagaacccggctga 438

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match	Length	ID		
1	438	100.0	438	21	AAZ49727	Streptomyces grise
2	407	92.9	407	21	AAZ49728	Streptomyces grise
3	298.2	68.1	407	21	AAZ49731	Streptomyces netro
4	259.8	59.3	407	21	AAZ49730	Streptomyces golde
5	239	54.6	407	21	AAZ49729	Streptomyces albus
6	77.2	17.6	45624	22	AAF88315	S. spinosa DNA fra
7	77.2	17.6	50000	22	AAF88312	S. spinosa DNA fra
8	58.6	13.4	438	21	AAZ49727	Streptomyces grise
9	55.2	12.6	65140	22	AAD17184	Streptomyces hours

10	55.2	12.6	125401	22	AAD17186	Streptomyces hours
11	54.8	12.5	80161	20	AAZ21501	DNA fragment of Sa
12	54.4	12.4	1549	16	AAQ88151	Xanthomonas L-glut
13	53.6	12.2	407	21	AAZ49728	Streptomyces grise
14	53.4	12.2	1313	22	AAF74867	Leishmania major p
15	53.4	12.2	1981	20	AAZ37293	Pseudomonas alpha-
16	53.4	12.2	1981	20	AAZ37294	Pseudomonas alpha-
17	53.4	12.2	1981	20	AAZ37295	Pseudomonas alpha-
18	53.4	12.2	1981	20	AAZ37296	Pseudomonas alpha-
19	53.4	12.2	1981	20	AAZ37297	Pseudomonas alpha-
20	53.4	12.2	1981	20	AAZ37298	Pseudomonas alpha-
21	53.4	12.2	3496	12	AAQ14465	Maltopentaose synt
22	52.6	12.0	154746	24	AAD25519	Human herpesvirus
23	51	11.6	117213	19	AAV62176	HSV-2 strain SB5 C
24	50.4	11.5	42000	21	AAAG63349	Streptomyces globi
25	50.4	11.5	63164	21	AAAG63348	Streptomyces globi
26	50.4	11.5	109519	22	AAQ08693	Micromonospora DNA
27	49.6	11.3	2900	13	AAQ20372	P. glumae PGI lipas
28	49.6	11.3	58857	21	AAA58471	Nucleotide sequenc
29	49.6	11.3	77536	21	AAA14651	Nucleotide sequenc
30	49	11.2	29879	14	AAQ46806	eryA region of S.
31	48.8	11.1	16767	22	AAF88339	S. spinosa DNA fra
32	48.8	11.1	50000	22	AAF88313	S. spinosa DNA fra
33	48.8	11.1	50000	22	AAF88316	S. spinosa DNA fra
34	48.8	11.1	109519	22	AAQ08693	Micromonospora DNA
35	48.6	11.1	3162	23	AAAS1499	Pseudomonas aerugi
36	48	11.0	1104	22	AAAS14801	DNA encoding Sebek
37	48	11.0	2064	14	AAQ52638	Streptomyces fradi
38	47.4	10.8	1161	15	AAQ53295	Consensus sequence
39	47.2	10.8	2488	22	AAH27511	Pseudomonas ATC hy
40	47.2	10.8	125401	22	AAD17186	Streptomyces hours
41	46.6	10.6	27541	22	AAD17185	Streptomyces hours
42	46.4	10.6	1542	21	AAZ60717	DNA encoding a bet
43	46.4	10.6	1542	22	AAD14504	S. clavuligerus cl
44	46.4	10.6	11604	22	AAD14501	Streptomyces clavu
45	46.4	10.6	15079	16	AAQ91580	S. clavuligerus cl

ALIGNMENTS

RESULT 1

AAZ49727	AAZ49727	standard; DNA; 438 BP.
ID	AAZ49727	standard; DNA; 438 BP.
XX	AAZ49727;	
AC	AAZ49727;	
XX	18-APR-2000	(first entry)
DT	18-APR-2000	(first entry)
XX	Streptomyces griseus	ssgA gene-1.
DE	Streptomyces griseus	ssgA gene-1.
XX	ssgA; liquid culture; filamentous bacteria; secondary metabolite;	
KW	mycelium; antibiotic, antitumour agent; immunosuppressive agent;	
KW	hypocholesterolemic agent; enzyme inhibitor; antimigraine agent;	
KW	herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23345;	
KW	bioinsecticide; receptor agonist; antagonist; biomass; ds.	
XX	Streptomyces griseus.	
OS	Streptomyces griseus.	
XX	Key	Location/Qualifiers
FT	CDS	31...438
FT		/*tag= a
FT		/product= "SsgA protein"
XX	WO200000613-A1.	
PN	WO200000613-A1.	
XX	06-JAN-2000.	
PD	06-JAN-2000.	
XX	25-JUN-1999;	99WO-NL00395.
PF	25-JUN-1999;	99WO-NL00395.
XX	26-JUN-1998;	98EP-0202148.
PR	26-JUN-1998;	98EP-0202148.
XX	(OYLE-) RIJKSUNIV LEIDEN.	
PA	(OYLE-) RIJKSUNIV LEIDEN.	

PA (NEWP-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.  
 XX Van Wezel GP, Kraal B, Luiten RGM;  
 XX WPI: 2000-147269/13.  
 DR P-PSDB; AAY44649.  
 XX  
 XX Reducing branching and enhancing fragmentation in filamentous  
 PT microorganisms used to improve their liquid culturing properties -  
 XX  
 PS Claim 1; Page 27; 60pp; English.  
 XX  
 XX The present sequence is a DNA (ATCC 23345) encoding S. griseus SsgA  
 CC protein. SsgA reduces branching and fragment septation and enhances  
 CC fragmentation of mycelium in liquid culture resulting in lower viscosity  
 CC of culture broths. Filamentous bacteria can be transformed with ssgA  
 CC gene-containing plasmid to enhance the production of secondary  
 CC metabolites such as, antibiotics, antitumour agents, immunosuppressive  
 CC agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine  
 CC agents, herbicides, antiparasitic agents, ruminant growth promoters,  
 CC bioinsecticides, receptor agonists and antagonists and biomass.  
 XX  
 SQ Sequence 438 BP; 68 A; 145 C; 153 G; 72 T; 0 other;

Query Match 100.0%; Score 438; DB 21; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-75;  
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcgcgagtcggttcaagcagaggttcagatgagcttctctccgagagctctcg 60  
 Db 1 atcgcgagtcggttcaagcagaggttcagatgagcttctctccgagagctctcg 60

QY 61 ttccgtattccgttgagctccgatacagaggtcgcgatccgtatgccatccgagatgac 120  
 Db 61 ttccgtattccgttgagctccgatacagaggtcgcgatccgtatgccatccgagatgac 120

QY 121 ttccaccttcccgcgatgcccctgacctggcggttgcggttcgcccgcgagctgctggac 180  
 Db 121 ttccaccttcccgcgatgcccctgacctggcggttgcggttcgcccgcgagctgctggac 180

QY 181 gggctcaacagccgagcggcgagcgatgtcacatcgcccgaccgagccgagggc 240  
 Db 181 gggctcaacagccgagcggcgagcgatgtcacatcgcccgaccgagccgagggc 240

QY 241 ctgcgagatgtccacatcccggtccaggtcgcgagcagcgtgcgtgttccggcgggg 300  
 Db 241 ctgcgagatgtccacatcccggtccaggtcgcgagcagcgtgcgtgttccggcgggg 300

QY 301 acggcaccgctggtggttctctcgacggagcagcaagctcgtgcgctcggccaggag 360  
 Db 301 acggcaccgctggtggttctctcgacggagcagcaagctcgtgcgctcggccaggag 360

QY 361 cacacgtggtgacttcgacggcaactgagagcagcactggcgccatccctgcgcag 420  
 Db 361 cacacgtggtgacttcgacggcaactgagagcagcactggcgccatccctgcgcag 420

QY 421 gacgagaacccggtga 438  
 Db 421 gacgagaacccggtga 438

RESULT 2  
 AAZ49728  
 ID AAZ49728 standard; DNA; 407 BP.  
 XX  
 AC AAZ49728;  
 XX  
 XX 18-APR-2000 (first entry)  
 XX Streptomyces griseus ssgA gene-2.  
 XX  
 KW ssgA; liquid culture; filamentous bacteria; secondary metabolite;

KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
 KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
 KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23345;  
 XX bioinsecticide; receptor agonist; antagonist; biomass; ds.  
 OS Streptomyces griseus.  
 XX  
 XX Key Location/Qualifiers  
 FH mat\_peptide 1..405  
 FT /\*tag= a  
 FT /label= SsgA\_protein  
 XX  
 XX WO200000613-A1.  
 XX  
 PD 06-JAN-2000.  
 XX  
 PF 25-JUN-1999; 99WO-NL00395.  
 XX  
 PR 26-JUN-1998; 98EP-0202148.  
 XX  
 XX (UYLE-) RIJKSUNIV LEIDEN.  
 PA (NEWP-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.  
 XX  
 XX Van Wezel GP, Kraal B, Luiten RGM;  
 XX  
 DR WPI: 2000-147269/13.  
 DR P-PSDB; AAY44649.  
 XX  
 XX Reducing branching and enhancing fragmentation in filamentous  
 PT microorganisms used to improve their liquid culturing properties -  
 XX  
 PS Disclosure; Fig 5; 60pp; English.  
 XX  
 XX The present sequence is a DNA (ATCC 23345) encoding S. griseus SsgA  
 CC protein. SsgA reduces branching and fragment septation and enhances  
 CC fragmentation of mycelium in liquid culture resulting in lower viscosity  
 CC of culture broths. Filamentous bacteria can be transformed with ssgA  
 CC gene-containing plasmid to enhance the production of secondary  
 CC metabolites such as, antibiotics, antitumour agents, immunosuppressive  
 CC agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine  
 CC agents, herbicides, antiparasitic agents, ruminant growth promoters,  
 CC bioinsecticides, receptor agonists and antagonists and biomass.  
 XX  
 XX Sequence 407 BP; 60 A; 139 C; 142 G; 66 T; 0 other;

Query Match 92.9%; Score 407; DB 21; Length 407;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-69;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 atgagcttctctctccgagagctctctccgtattccgttattccgttgagctccgatacag 90  
 Db 1 atgagcttctctctccgagagctctctccgtattccgttattccgttgagctccgatacag 90

QY 91 gtcgggagctcgatgccatcccgatgacgttccacattcccgcgatgccctgtgacc 150  
 Db 61 gtcgggagctcgatgccatcccgatgacgttccacattcccgcgatgccctgtgacc 120

QY 151 tggcgcttcggccgcgagctgctgtagcggtcacaagcccgagcggcgagcggcgat 210  
 Db 121 tggcgcttcggccgcgagctgctgtagcggtcacaagcccgagcggcgagcggcgat 180

QY 211 gtgcacatcgcccgaccgagccgagggcctcgaggatgtccacatccggctccaggtc 270  
 Db 181 gtgcacatcgcccgaccgagccgagggcctcgaggatgtccacatccggctccaggtc 240

QY 271 ggcgggagcctgctgctgttccgggggggagcagcagcctggtggcttccctgcacgg 330  
 Db 241 ggcgggagcctgctgctgttccgggggggagcagcagcctggtggcttccctgcacgg 300

QY 331 acggacaagctgctgcgctcgcgagagcacacgctgggtgacttcgacgggcaactg 390  
 Db 301 acggacaagctgctgcgctcgcgagagcacacgctgggtgacttcgacgggcaactg 360









Db 24645 aacgttagcagcagcggtcaggtgactgggtgatccgcgcgacgtgctggccgacggg 24704  
QY 184 ctcaacagccgagcgcgagcgagcgatgtgcacatcgccgagccgagccgagggcctc 243  
Db 24705 ctgatccgagggcagcgagcgatgtgcggtatcgccc---ctcagcgggttttccg 24761  
QY 244 ggagatgtccacatcccgctccaggtcgccgagcgtgcgctgttccgaggggagcg 303  
Db 24762 gggtgtgtgacgagatgagctcgccgtcgggcagcgccctctctcgaggtgaatgct 24821  
QY 304 gcacgcgtgtgtggtctctctcgcgagcagcgacagctgtgcgctcgccgagagcagc 363  
Db 24822 gaccagcttgcgacttcttgaacgacacctacgctggtgcgaacctgtgtgataaac 24881  
QY 364 acgtggtgacttcgacgcaacctgagagcagcactgggcccgcacatcct 413  
Db 24882 cgttgatgaacgtcgacgaggtgctgagccagctgctctcgcaaacct 24931

## RESULT 8

AAZ49727/C

ID AAZ49727 standard; DNA; 438 BP.

XX

AC AAZ49727;

XX

XX 18-APR-2000 (first entry)

XX

XX Streptomyces griseus ssga gene-1.

XX

KW ssga; liquid culture; filamentous bacteria; secondary metabolite;

KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;

KW hypcholesterolaemic agent; enzyme inhibitor; antimigraine agent;

KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23345;

KW bioinsecticide; receptor agonist; antagonist; biomass; ds.

XX

OS Streptomyces griseus.

XX

XX Key Location/Qualifiers

FH CDS

FT 31..438

FT /\*tag= a

FT /product= "Ssga protein"

XX

XX WC200000613-A1.

XX

XX 06-JAN-2000.

XX

XX 25-JUN-1999; 99WO-NL00395.

XX

XX 26-JUN-1998; 98EP-0202148.

XX

XX (UYLE-) RIJKSUNIV LEIDEN.

XX

XX (NEW-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.

XX

XX Van Wezel GP, Kraal B, Luiten RGM;

XX

XX WPI; 2000-147269/13.

XX

XX P-PSDB; AAY44649.

XX

XX Reducing branching and enhancing fragmentation in filamentous

XX

XX microorganisms used to improve their liquid culturing properties -

XX

XX Claim 1; Page 27; 60pp; English.

XX

XX The present sequence is a DNA (ATCC 23345) encoding S. griseus Ssga

XX protein. Ssga reduces branching and fragment septation and enhances

XX fragmentation of mycelium in liquid culture resulting in lower viscosity

XX of culture broths. Filamentous bacteria can be transformed with ssga

XX gene-containing plasmid to enhance the production of secondary

XX metabolites such as, antibiotics, antitumour agents, immunosuppressive

XX agents, hypcholesterolaemic agents, enzyme inhibitors, antimigraine

XX agents, herbicides, antiparasitic agents, ruminant growth promoters,

XX bioinsecticides, receptor agonists and antagonists and biomass.

xx  
SQ Sequence 438 BP; 68 A; 145 C; 153 G; 72 T; 0 other;

Query Match 13.4%; Score 58.6; DB 21; Length 438;  
Best Local Similarity 47.2%; Pred. No. 0.0075;  
Matches 178; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 12 ggttcaagcagaggtcatgatgagcttcctcgctcctcgagagagctcgttccgtattcc 71

Db 388 GGTTCGCGTCAAGTCACCCAGCGTGTCTCTCGCGAGCGGACGAGCTTGTCCGTCC 329

QY 72 ggttgagctccgaacagaggtcgcgatccgtatgcatccgagatgacgttccaccctcc 131

Db 328 GGTGAGGAAAGCCACACGCGGTGCGCTCCCGCCCGGAACAGCGCGGTCCCGCGCGA 269

QY 132 cggcagtcgcccctgtgacctggcgcttcgcccgcgagctgctgtgacggggtcacaag 191

Db 268 CCTGGAGCGGATGTGGACATCTCCGAGGCCCTCGGCTCGGTCCGCGCGCATGTGTCACAT 209

QY 192 cccgagcgcgacggcgatgtgcacatcgccgacccgagccgagggcctcgagatgt 251

Db 208 CGCGCTCGCGCTCGCGGAGGTGGAACGTATCCCGATGGCATACGGATCGCGGACCT 149

QY 252 ccacatccggctccaggtcgcgacgtgcgctgttcggcggggagcggccacgcgt 311

Db 148 TCACAGGGGCATCGCCGGGAAGGTGGAACGTATCCCGATGGCATACGGATCGCGGACCT 89

QY 312 ggtggcgttctcgcgacggcgagacagctcgtgcgcctcgcccgagagagcacacgcgtg 371

Db 88 CGTATCGGAGCTCCACCGGAATACGGAACGAGAGCTCTCGGAGACGAGGAAGTCAATCA 29

QY 372 tgacttcgacggcaacc 388

Db 28 TGACCTCTGCTTGAACC 12

## RESULT 9

AAD17184

ID AAD17184 standard; DNA; 65140 BP.

XX

AC AAD17184;

XX

XX 29-NOV-2001 (first entry)

XX

XX Streptomyces noursei nysl DNA of nystatin PKS gene cluster.

XX

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

KW antifungal; antibiotic; nysI; ds.

XX

OS Streptomyces noursei.

XX

FH Key Location/Qualifiers

FT CDS

FT complement (1..1035)

FT /\*tag= a

FT /product= "NysD2 partial protein"

FT /note= "CDS does not include stop codon"

FT CDS

FT complement (1036..2576)

FT /\*tag= b

FT /product= "NysD1 protein"

FT 2806..6906

FT /\*tag= c

FT /product= "NysA protein"

FT 6952..16530

FT /\*tag= d

FT /product= "NysB protein"

FT 16550..49840

FT /\*tag= e

FT /product= "NysC protein"

FT 50260..51015

FT /\*tag= f

FT /product= "NysE protein"

FT 51405..54305

FT CDS

```

FT FT /*tag= g
FT FT /product= "NysR1 protein"
FT FT 54329..57190
FT FT /*tag= h
FT FT /product= "NysR2 protein"
FT FT /note= "CDS does not include start codon"
FT FT 57180..59963
FT FT /*tag= i
FT FT /product= "NysR3 protein"
FT FT 60415..61047
FT FT /*tag= j
FT FT /product= "NysR4 (short) protein"
FT FT /note= "CDS does not include start codon"
FT FT 61736..62497
FT FT /*tag= k
FT FT /product= "NysR5 protein"
FT FT /note= "CDS does not include start codon"
FT FT complement (62551..63615)
FT FT /*tag= l
FT FT /product= "ORF2 protein"
FT FT /note= "CDS does not include start codon"
FT FT 63765..64961
FT FT /*tag= m
FT FT /product= "ORF1 protein"
FT FT
XX XX WO200159126-A2.
XX XX 16-AUG-2001.
XX XX
XX XX 08-FEB-2001; 2001WO-GB00509.
XX XX
XX XX 08-FEB-2000; 2000GB-0002840.
XX XX 10-APR-2000; 2000GB-0008786.
XX XX 14-APR-2000; 2000GB-0009387.
XX XX
XX XX (UYNO-) UNIV NORGE TEKNIKSKAPPELIGE.
XX XX (SNTF) SIMTEF STIFTELSEN IND TEK FORSK.
XX XX PA (ALPH-) ALPHARMA AS.
XX XX PA (SINV-) SINVENT AS.
XX XX PA (DZIE/) DZIEGLEWSKA H.
XX XX PA (ZOTC/) ZOTCHEV S B.
XX XX PA (SEKU/) SEKUROVA O N.
XX XX PA (FJAE/) FJAEVRIK E.
XX XX PA (BRAU/) BRAUTASET T.
XX XX PA (STRO/) STROM A R.
XX XX
XX XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX XX
XX XX WPI; 2001-557614/62.
XX XX P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
XX XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX XX
XX XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX XX useful as antibiotics and antifungals -
XX XX
XX XX Claim 2; Page 116-151; 266pp; English.
XX XX
XX XX The present invention relates to the cloning and sequencing of the gene
XX XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX XX involved in the biosynthesis of the macrolide antibiotic nystatin.
XX XX The nystatin PKS is useful as antifungal antibiotics. The present
XX XX sequence is a Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
XX XX
XX XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
XX XX
XX XX Query Match 12.6%; Score 55.2; DB 22; Length 65140;
XX XX Best Local Similarity 49.3%; Pred. No. 0.03;
XX XX Matches 201; Conservative 0; Mismatches 203; Indels 4; Gaps 2;
XX XX
XX XX 23 aggtcatgatgacttcctcgtccgagagcttcgttcctcgtattccggtggagctcc 82
XX XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX XX

```

```

Db 50786 aggcacatcgagacactacgtgccgcgccgacgtcacccgtccgagcgccgtgaccgtcc 50845
Qy 83 gatacaggtcgcgacatccgtatgccatcccgatgacgttccacattcccgcgatgcc 142
Db 50846 tcaccgagcgcgaccccgagacacctccctggacgagggcgga---ggcgtggcgcgcc 50902
Qy 143 ctgtgacctggcgttcgcccgcgagctgctgctggacgggctcaacagcccgagcgcg 202
Db 50903 acaccacggggacttcgacctcaagtgcttcccggtggcactcttcgtcagctccg 50962
Qy 203 acggcgatgtcacatcggcccgcgagccgagccgagggcctcgagagatgtcacatccggc 262
Db 50963 aggcgccgcatcatgctgtctccgggcgacactcgccgacgctagcgggcgcc 51022
Qy 263 tccaggtcgcgcgagaccgtgcgtgttccggcgagggagcgcaccgcctgtggcgcttc 322
Db 51023 actcggcagggccggtgcgtgtctccgacacggcccgcgctgagacgggacac 51082
Qy 323 tcgaccgagcgacaaagctcgtgcgctcggc-caggagacacacgctgggtgacttcgac 381
Db 51083 catgccaccgagcgcgctgtcgtgcgctgtggagacccgtcccgctgcgtacgc 51142
Qy 382 ggcacactggagggacgacactggcgcatctcgcgagggagagagaac 429
Db 51143 gggcgcgaggacgtgctcggcgccgaggtcctggagagagacc 51190

RESULT 10
AAAD17186
ID AAD17186 standard; DNA; 125401 BP.
XX AC
XX AC AAD17186;
XX DT 29-NOV-2001 (first entry)
XX DE Streptomyces noursei nystatin PKS gene cluster DNA.
XX KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX KW antifungal; antibiotic; ds.
XX OS Streptomyces noursei.
XX FH Key Location/Qualifiers
XX FT 6337..34771
XX FT CDS /*tag= a
XX FT /product= "NysI complete protein"
XX FT 34792..51099
XX FT CDS /*tag= b
XX FT /product= "NysJ protein"
XX FT 51155..57355
XX FT CDS /*tag= c
XX FT /product= "NysK protein"
XX FT 57503..58687
XX FT CDS /*tag= d
XX FT /product= "NysL protein"
XX FT complement (58786..58980)
XX FT CDS /*tag= e
XX FT /product= "NysM protein"
XX FT /note= "CDS does not include start codon"
XX FT complement (59045..60241)
XX FT CDS /*tag= f
XX FT /product= "NysN protein"
XX FT /note= "CDS does not include start codon"
XX FT complement (60238..61296)
XX FT CDS /*tag= g
XX FT /product= "NysD2 complete protein"
XX FT 120628..121308
XX FT CDS /*tag= h
XX FT /product= "NysR4 (long) protein"
XX PN WO200159126-A2.
XX PD 16-AUG-2001.

```

```
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX
XX 10-APR-2000; 2000GB-0008786.
XX
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX
XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
XX
XX (ALPH-) ALPHARMA AS.
XX
XX (SINV-) SINVENT AS.
XX
XX (DZIE/) DZIEGLEWSKA H.
XX
XX (ZOTC/) ZOTCHEV S B.
XX
XX (SEKU/) SEKUROVA O N.
XX
XX (FJAE/) FJAERVIK E.
XX
XX (BRAU/) BRAUTASET T.
XX
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI: 2001-557614/62.
XX
XX P-PSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX AAE10149, AAE10150.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
XX
XX Claim 1; Page 188-254; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
XX
XX
XX Query Match 12.68; Score 55.2; DB 22; Length 125401;
XX Best Local Similarity 49.38; Pred. No. 0.029;
XX Matches 201; Conservative 0; Mismatches 203; Indels 4; Gaps 2;
XX
XX Qy 23 aggtcatgatgagctctcgtccgagagagagctctcgtccgtattccgtgtgagctcc 82
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 111047 aggcacatgagacctaccgtctccgcccagctcacgctccggtcgccgtgacctcc 111106
XX
XX Qy 83 gatacagggtcggtccgtatgccatccgagatgacgttccacctcccggtgagctcc 142
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 111107 tcaccggcagccgcagcccaagacctccctggagcagggccga---ggcgtggcgcggcc 111163
XX
XX Qy 143 ctgtgacctggcgttcgcccgcagctgctgtcgtgagcaggtcacaacagccgagcggcg 202
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 111164 acaccacggggagcttcgacctcaagtgcttcccggtgggacattcttcgtcagctccg 111223
XX
XX Qy 203 acgctgcatgtgcacatcgcccgaccgagcccgagggcctcggagatgtccacatccggc 262
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 111224 agccccggcgatcatgactgctcgtcggcgacacctcgcgcgcaacgtagcggcgcc 111283
XX
XX Qy 263 tccaggttcggcgagacccgtgcgtgttcggggcgggagcagccgctggtggtcc 322
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 111284 actcggcagcggcggtgctgcgtgtcgcgcagcggcccccgccgctgagacggcac 111343
XX
XX Qy 323 tcgacggagcagcaaatcgttcgctcgtcgcgcagcagcagcagcagcagcagcagcagc 381
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 111344 catgccaccggcgagcgcgctgctgctgctgctgctgctgctgctgctgctgctgctgct 111403
XX
XX Qy 382 ggcaacctgagggagcagcactggcgccgcatctcctccgagagagagagagac 429
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 111404 gggcgcgagagacgtgctgcggccgagcagcagcagcagcagcagcagcagcagcagc 111451
XX
XX RESULT 11
```

```
FT /product= spnH
FT /note= "Spinosyn biosynthesis protein H"
FT complement (18541..19713)
FT /*tag= o
FT /product= spnG
FT /note= "Spinosyn biosynthesis protein G"
FT 20168..20995
FT /*tag= p
FT /product= spnF
FT /note= "Spinosyn biosynthesis protein F"
FT 21111..28898
FT /*tag= q
FT /product= spnA
FT /note= "Spinosyn biosynthesis protein A"
FT /trans_except= (Pos:26940..26942, aa:Pro)
FT 21126..22379
FT /*tag= r
FT /note= "Beta-ketosynthase domain (KS1): part of the
FT initiator module"
FT 22692..23669
FT /*tag= s
FT /note= "Acyl transferase domain (AT1): part of the
FT initiator module"
FT 23793..24041
FT /note= "Acyl carrier protein domain (ACP1): part of the
FT initiator module"
FT 24102..25349
FT /*tag= u
FT /note= "Beta-ketosynthase domain (KS1): part of
FT extender module 1"
FT 25683..26684
FT /*tag= v
FT /note= "Acyl transferase domain (AT1): part of
FT extender module 1"
FT 27582..28121
FT /*tag= w
FT /note= "Beta-ketoreductase domain (KR1): part of
FT extender module 1"
FT 28404..28649
FT /*tag= x
FT /note= "Acyl carrier protein domain (ACP1): part of
FT extender module 1"
FT 38916..35374
FT /*tag= y
FT /product= spnC
FT /note= "Spinosyn biosynthesis protein C"
FT 29024..30295
FT /*tag= z
FT /note= "Beta-ketosynthase domain (KS2): part of
FT extender module 2"
FT 30629..31621
FT /*tag= aa
FT /note= "Acyl transferase domain (AT2): part of
FT extender module 2"
FT 31697..32254
FT /*tag= ab
FT /note= "Dehydratase domain (DH2): part of extender
FT module 2"
FT 33035..34072
FT /*tag= ac
FT /note= "Enoyl reductase domain (ER2): part of
FT extender module 2"
FT 34082..34621
FT /*tag= ad
FT /note= "Beta-ketoreductase domain (KR2): part of
FT extender module 2"
FT 34886..30295
FT /*tag= ae
FT /note= "Acyl carrier protein domain (ACP2): part of
FT extender module 2"
FT 35419..44931
FT /*tag= af
```

```
FT /product= spnC
FT /note= "Spinosyn biosynthesis protein C"
FT 35518..36786
FT /*tag= ag
FT /note= "Beta-ketosynthase domain (KS3): part of
FT extender module 3"
FT 37108..38097
FT /*tag= ah
FT /note= "Acyl transferase domain (AT3): part of
FT extender module 3"
FT 38992..39528
FT /*tag= ai
FT /note= "Beta-ketoreductase domain (KR3): part of
FT extender module 3"
FT 39790..40035
FT /*tag= aj
FT /note= "Acyl carrier protein domain (ACP3): part of
FT extender module 3"
FT 40102..41373
FT /*tag= ak
FT /note= "Beta-ketosynthase domain (KS4): part of
FT extender module 4"
FT 41713..42705
FT /*tag= al
FT /note= "Acyl transferase domain (AT4): part of
FT extender module 4"
FT 43615..44157
FT /*tag= am
FT /note= "Beta-ketoreductase domain (KR4): part of
FT extender module 4"
FT 44431..44676
FT /*tag= an
FT /note= "Acyl carrier protein domain (ACP4): part of
FT extender module 4"
FT 44966..59752
FT /*tag= ao
FT /product= spnD
FT /note= "Spinosyn biosynthesis protein D"
FT 45077..46348
FT /*tag= ap
FT /note= "Beta-ketosynthase domain (KS5): part of
FT extender module 5"
FT 46691..47674
FT /*tag= aq
FT /note= "Acyl transferase domain (AT5): part of
FT extender module 5"
FT 47753..48310
FT /*tag= ar
FT /note= "Dehydratase domain (DH5): part of extender
FT module 5"
FT 49226..49771
FT /*tag= as
FT /note= "Beta-ketoreductase domain (KR5): part of
FT extender module 5"
FT 50009..50254
FT /*tag= at
```

Query Match 12.58; Score 54.8; DB 20; Length 80161;  
Best Local Similarity 52.1%; Pred. No. 0.035;  
Matches 147; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

```
Qy 109 atccggtgacgttccacccctcccgcgatgccctgtgacctggcggttcgcccgcgag 168
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 atctccatgaagctcaacgttaggcacgcgaggtcaggtggatggtgacgcccgcgac 61
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 169 ctgctgtgacggggtctcaacagcccgagcgcgacggtgtgcacatggtgcacgcccgcgac 228
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 ctgctggccgacgggtgatcgccgagcgagcgagcgatgtgcggtatcgcc---ct 118
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 229 gagcccgaggcctcggagatgtccacatccggtccaggtcggcgcgagaccgtgcgtg 288
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 cgacggggtttccgggggttggtcgtgagatgagctgcggtccggtcgggcaggtctcc 178
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 289 ttccgggggggacggcaccgctggtggtctctcgcacggagcgaacgctgctgcg 348
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

Db 179 ttgaggtgaatgctgaccagcttggtgacttcttgaacgacacctacgactggtcgaa 238
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 349 ctgagccagagcacacgctgggtgacttcgacgagcgaacctg 390
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 cctggtgatgaacacgctggtgatgaacgctgacgaggtgctg 280
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AAQ88151
ID AAQ88151 standard; DNA; 1549 BP.
XX
AC AAQ88151;
XX
DT 02-NOV-1995 (first entry)
XX
DE Xanthomonas L-glutamyl tRNA reductase.
XX
KW Xanthomonas campestris; L-glutamyl tRNA reductase; vitamin B12;
KW glutamic acid-1-semialdehyde; 5-aminolevulinic acid (ALA); haem;
KW chlorophyll; phycobillin; recombinant production; ds.
XX
OS Xanthomonas campestris.
XX
FH Key Location/Qualifiers
FT CDS 212..1492
FT /*tag= a
XX
PN JP07031480-A.
XX
PD 03-FEB-1995.
XX
PF 27-JUL-1993; 93JP-0184709.
XX
PR 27-JUL-1993; 93JP-0184709.
XX
PA (COSM-) COSMO SOGO KENKYUSHO KK.
XX
WPI; 1995-109535/15.
DR P-PSDB; AAR74619.
XX
XX DNA fragment encoding L-glutamyl tRNA reductase - derived from
PT Xanthomonas sp., useful for production of 5-aminolevulinic acid,
PT vitamin-B12, chlorophyll, etc.
XX
PS Claim 3; Pages 7-8; 9pp; Japanese.
XX
AC AAQ88151 encodes AAR74619 Xanthomonas campestris L-glutamyl tRNA
CC reductase. The DNA as part of an expression vector can be used to
CC transform a host cell, for the recombinant prodn. of the protein.
CC The protein can be used in the prodn. of vitamin B12, glutamic
CC acid-1-semialdehyde, 5-aminolevulinic acid, haem, chlorophyll and
CC phycobillin.
XX
SQ Sequence 1549 BP; 291 A; 516 C; 514 G; 228 T; 0 other;

Query Match 12.4%; Score 54.4; DB 16; Length 1549;
Best Local Similarity 53.2%; Pred. No. 0.046;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 218 tcgagccagacgagccgagggcctcgagatgtccacatccggtccaggtcgcgcg 277
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 tggacctgcgcgaacgcgcgcgttcgaggtgatgctgcgcgcgcgcgcgaatcgc 314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 278 accgtgcgtgttcggcgggggaacgacccgctggtggttcctcgaccgagcaga 337
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 tgcgtgcgtgtcccccgggtaagcgagggcgcgctgctgtccacctgaacctgacgagt 374
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 338 agctgcgtgcgtcgccgagagcagacacgctgggtgacttcgacgcaacctggaggagc 397
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 tgtatcgatggcgagggcgacacagcagctggttcacctggtggaacccacgcccgcg 434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Qy 398 cactggcgccatctctcgccgagagcagacgaacgccc 433
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 ctctgagcggttatctgtaccagcaccaggaagccc 470
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AAZ49728/c
ID AAZ49728 standard; DNA; 407 BP.
XX
AC AAZ49728;
XX
DT 18-APR-2000 (first entry)
XX
DE Streptomyces griseus ssgA gene-2.
XX
KW ssgA; liquid culture; filamentous bacteria; secondary metabolite;
KW mycelium; antibiotic; antitumour agent; immunosuppressive agent;
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;
KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23345;
KW bioinsecticide; receptor agonist; antagonist; biomass; ds.
XX
OS Streptomyces griseus.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..405
FT /*tag= a
FT /*label= SsgA_protein
XX
PN WO2000006113-A1.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-NL00395.
XX
PR 26-JUN-1998; 98EP-0202148.
XX
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZ.
XX
PI Van Wezel GP, Kraal B, Luiten RGM;
XX
WPI; 2000-147269/13.
DR P-PSDB; AAY44649.
XX
PT Reducing branching and enhancing fragmentation in filamentous
PT microorganisms used to improve their liquid culturing properties -
XX Disclosure; Fig 5; 60pp; English.
XX
CC The present sequence is a DNA (ATCC 23345) encoding S. griseus SsgA
CC protein. SsgA reduces branching and fragment septation and enhances
CC fragmentation of mycelium in liquid culture resulting in lower viscosity
CC of culture broths. Filamentous bacteria can be transformed with ssgA
CC gene-containing plasmid to enhance the production of secondary
CC metabolites such as, antibiotics, antitumour agents, immunosuppressive
CC agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine
CC agents, herbicides, antiparasitic agents, ruminant growth promoters,
CC bioinsecticides, receptor agonists and antagonists and biomass.
XX
SQ Sequence 407 BP; 60 A; 139 C; 142 G; 66 T; 0 other;

Query Match 12.2%; Score 53.6; DB 21; Length 407;
Best Local Similarity 46.9%; Pred. No. 0.068;
Matches 167; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

Qy 12 ggttaacagagggatcatgatgacttcctcgtctccgagagcgtctcgttccgtattcc 71
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 GGTTCGCGTCGAAGTCAACCCAGCGTGTCTCTCGCGAGCGCACGAGCTTGTCCGTCC 299
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 72 ggtgagctccgatcacgaggtcgccgatcgtatgcattccggtacgcttcacacctcc 131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 GGTGAGGAGAACGCCACACAGCGGTGCGGTCCCGCCCGGAGACAGCGTCCGCGCCGA 239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



Qy	132	cgcgcatgccctgtgacctggcgcttcgcgcgcgcagctgctcgtggacggcgctcaaacag	191
Db	238	CCTCGGACCGGGATGTGGACATCTCCGAGGCCCTCGGGCTCGGTTCGGCCGATGTGCACAT	179
Qy	192	cccgagcgcgcgcgcgatgtgcacatcggccgaccgagcccgaaggcctcggagatgt	251
Db	178	CGCGTTCGGCTCGGGCTGTTCAGCCCCGTCCAGACGACTCGCGGCCGAACGCCCAAG	119
Qy	252	ccacatcccgctcacaggtggcgcgcgcacgctgcgctgttcccgggcgggacgcgcaccgc	311
Db	118	TACAGGGGGCATCGCCGGGAAGTGGAACGTCATCCGGATGCGATACGATCGCCGACCT	59
Qy	312	ggtggcgcttcctgcgcgcgcgcgcagaagtctgcgcctcgcgcgcgcgcgcgcgcgcgc	367
Db	58	CGTATCGGAGCTCCACCGAATACGGAACGAGAGCTCCTCGGAGACGAGGAAGCTC	3
 RESULT 14 AAF74867/c			
ID	AAF74867 standard; DNA; 1313 BP.		
XX	AAF74867;		
XX			
DT	22-MAY-2001 (first entry)		
XX			
DE	Leishmania major PPG nucleotide sequence.		
XX			
KW	Human; hPPG-1; PPG; proteophosphoglycan; detection; ds.		
XX	Leishmania major.		
OS			
PX	CN1272542-A.		
PN			
PD	08-NOV-2000.		
XX			
PF	11-APR-2000; 2000CN-0115368.		
PR			
XX	11-APR-2000; 2000CN-0115368.		
(NANF-) NANFANG RES CENT STATE HUMAN GENE GROUP.			
PA			
XX			
PI	Li N, Xiao H, Kang B;		
XX			
DR	WPI; 2001-183595/19.		
DR	P-PSDB; AAB74609.		
XX			
PT	Human phosphoglycan protein and its coded sequence -		
XX			
PS	Example 2; Page 17-19; 22pp; Chinese.		
XX			
CC	The present invention describes a human proteophosphoglycan protein		
CC	designated hPPG-1. Also describes are methods for the preparation		
CC	and detection of hPPG-1. The present sequence encodes the Leishmania		
CC	major PPG protein which is used in comparison with the hPPG-1 protein		
CC	in an example from the present invention.		
XX			
SQ	Sequence 1313 BP; 194 A; 457 C; 367 G; 295 T; 0 other;		

D	b	419	GCAGACGACGAAGACGCCGATGGCGGGAGCTGCTGTGGACGGCGCAGCACGACGAGGAC	368
Q	y	199	ggcgcagcgcatgtgacatcgccgcgaccgagccgaggcctcggaatgtccacatc	258
D	b	359	GCCGACGCGCGGAGCTGTCTGGACGGCGCAGACGACGAAGACGCCGACGCGCGGAG	300
Q	y	259	cggctccagtgtcggcgagacctgcgctgttcccgggcggggacgcccgctgggtggcg	318
D	b	299	,CTGCTGTGGACGGCGCAGACGACGAAGACGCCGATGGCGCGGAGCTGTGCTGGACGCT	240
Q	y	319	tctctgaccgcgagcagaagtctgctgcgctcggccagagcacacgctgggtgaattc	378
D	b	239	GCAGACGACGAGGACGCCGACGGCGGAGCTGCTGTGGACGGCGCAGCACGACGACG	180
Q	y	379	gacggcaactggagcagcactgggcgcgatctcgccgagagcagaaacgcgcg	433
D	b	179	GACGCTGACGGCGGAGCTGTCTGTGGACGGCGCAGACGACGAAGACGCCG	125
RESULT 15				
AAAX37293				
I	D	AAAX37293 standard; DNA; 1981 BP.		
X	X			
A	C	AAAX37293;		
X	X			
D	T	27-JUL-1999 (first entry)		
X	X			
D	E	Pseudomonas alpha-amylase variant encoding DNA (FERM BP-6116).		
X	X			
K	W	Pseudomonas; maltopentaose; alpha-amylase; enzyme; variant; ss.		
O	S	Pseudomonas sp.		
O	S	Synthetic.		
X	X			
P	N	JF11123081-A.		
X	X			
P	D	11-MAY-1999.		
X	X			
P	F	21-OCT-1997; 97JP-0305071.		
X	X			
P	R	21-OCT-1997; 97JP-0305071.		
X	X	(NORQ ) NORINSUISANSHO SHOKUHI SGO.		
P	A	WPI; 1999-340517/29.		
X	X	P-PSDB; AAY17904.		
D	R			
D	R			
X	X			
P	T	New alpha-amylase gene - useful for production of malto-pentaose		
X	X			
P	S	Claim 4; Page 6-8; 23pp; Japanese.		
X	X			
C	C	Sequences AAAX37293-298 represent nucleic acids encoding alpha-amylase		
C	C	variants (AAAY17904-Y17909). The variants are derived from the		
C	C	Pseudomonas sp. KO-8940-derived maltopentaose-forming amylase (Shida,		
C	C	et al, Biosci. Biotech, Biochem. Vol. 56, 76-80, 1992 ). The variants		
C	C	have a replacement of the 57th or the 130th amino acid residue in the		
C	C	sequence as compared to the wild-type. The sequences are deposited un-		
C	C	der accession numbers FERM BP-6116, 6119, 6115, 6117, 6118 or 6114. The		
C	C	variants can be used to transform plants to produce maltopentaose. The		
C	C	enzymes prepared from the alpha-amylase variants produce maltopentaos		
C	C	in a high yield.		
X	X			
S	Q	Sequence 1981 BP; 342 A; 678 C; 678 G; 283 T; 0 other:		

	Query Match	12.2%	Score 53.4;	DB 20;	Length 1981;
	Best Local Similarity	51.0%;	Pred. No.	0.072;	
	Matches 126; Conservative	0;	Mismatches 121;	Indels 0;	Gaps 0;
Qy	147	gacctgggcgttcggcccgagctcgtctgtgagcgggctcaacgccgagcgcgacgg	206		
Db	582	gatctccgaactacctgfcscgacccttggtcccagaataggctcaagqctkgcgcgtcatgc	641		



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 08:23:30 ; Search time 165.56 Seconds  
(without alignments)  
649.840 Million cell updates/sec

Title: US-09-749-185-1

Perfect score: 438

Sequence: 1 atgcgcgagtcggttcaagc.....aggagcagaacgcgcgtga 438

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PCRUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	54.8	12.5	80161	3	US-09-036-987A-1
2	54.8	12.5	80161	4	US-09-370-700-1
3	53.4	12.2	1981	3	US-09-017-706-3
4	53.4	12.2	1981	3	US-09-017-706-4
5	53.4	12.2	1981	3	US-09-017-706-5
6	53.4	12.2	1981	3	US-09-017-706-6
7	53.4	12.2	1981	3	US-09-017-706-7
8	53.4	12.2	1981	3	US-09-017-706-8
9	49.6	11.3	2900	1	US-08-034-650-9
10	49.6	11.3	2900	1	US-08-449-015-9
11	49	11.2	1155	2	US-08-387-942C-7
12	49	11.2	11219	1	US-07-842-734C-1
13	49	11.2	11219	3	US-08-439-009A-1
14	48	11.0	2064	4	US-08-343-428-1
15	46.4	10.6	1542	4	US-09-385-028-16
16	46.4	10.6	11604	4	US-09-385-028-13
17	46.4	10.6	15079	4	US-09-385-028-1
18	46	10.5	2745	1	US-08-363-255-1
19	46	10.5	2745	1	US-08-363-255-13
20	45.8	10.5	1155	2	US-08-387-942C-18
21	45.8	10.5	12588	2	US-08-387-942C-1
22	45.4	10.4	2439	4	US-09-624-693A-18
23	44.8	10.2	1212	6	5212296-17
24	44.8	10.2	1998	6	5212296-8
25	44	10.0	8931	3	US-09-051-019-1
26	44	10.0	20235	1	US-07-642-734C-3
27	44	10.0	20235	3	US-08-439-009A-3

28	43.8	10.0	44377	2	US-08-804-227C-7	Sequence 7, Appli
29	43.8	10.0	44377	2	US-08-804-198-1	Sequence 1, Appli
30	43.6	10.0	1683	1	US-07-945-283-3	Sequence 3, Appli
31	43.6	10.0	1831	6	5215881-1	Patent No. 5215881
32	43.6	10.0	1831	6	5215881-3	Patent No. 5215881
33	43.6	10.0	8438	1	US-07-945-283-1	Sequence 1, Appli
34	43.4	9.9	9515	1	US-08-920-812-13	Sequence 13, Appli
35	43.4	9.9	9515	1	US-08-920-827-13	Sequence 13, Appli
36	43.4	9.9	9515	1	US-08-921-177-13	Sequence 13, Appli
37	43.4	9.9	9515	1	US-08-362-577C-13	Sequence 13, Appli
38	43.4	9.9	9515	2	US-08-920-828-13	Sequence 13, Appli
39	43	9.8	20235	1	US-07-642-734C-3	Sequence 3, Appli
40	43	9.8	20235	3	US-08-439-009A-3	Sequence 3, Appli
41	43	9.8	71989	4	US-09-443-501A-2	Sequence 2, Appli
42	42.4	9.7	1294	3	US-09-025-691-2	Sequence 7, Appli
43	42.4	9.7	44377	2	US-08-804-227C-7	Sequence 1, Appli
44	42.4	9.7	44377	2	US-08-804-198-1	Sequence 46, Appli
45	42.2	9.6	999	2	US-08-761-277A-46	

ALIGNMENTS

RESULT 1

US-09-036-987A-1

; Sequence 1, Application US/09036987A

; Patent No. 6143526

; GENERAL INFORMATION:

; APPLICANT: Baltz, Richard H.

; APPLICANT: Broughton, Mary C.

; APPLICANT: Crawford, Kathryn P.

; APPLICANT: Madduri, Krishnamurthy

; APPLICANT: Merlo, Donald J.

; APPLICANT: Treadway, Patti J.

; APPLICANT: Turner, Jan R.

; APPLICANT: Waldron, Clive

; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide

; TITLE OF INVENTION: Production

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dow Agrosciences LLC Patent Department

; STREET: 9330 Zionsville Road

; CITY: Indianapolis

; STATE: Indiana

; COUNTRY: USA

; ZIP: 46268

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/036,987A

; FILING DATE: 09-MAR-1998

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Stuart, Donald R

; REGISTRATION NUMBER: 28,479

; REFERENCE/DOCKET NUMBER: 50,608

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (317)337-4816

; TELEFAX: (317)337-4847

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 80161 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-09-036-987A-1

Query Match 12.5% Score 54.8; DB 3; Length 80161;



Qy 147 gacctgggcgttcggcgcgagctgctgctgacgggctcaacagccccgagcgacgg 206

Db	582	gatctccgactaacttggtcgacctggtccagatggcggtcgaaggcctgctgcggtcgatgc	641
Qy	207	cgatgtgcacatcgcccgaccgagcccgagggcttcgagatgtccacatccggctcca	266
Db	642	ggccaagcacatcagcccgaccgacctgggtgccatcatcgacagcgtcaacgcgcgcac	701
Qy	267	ggtcgcgcggaacgctgcgctgttcggcgcggggaacgcgcacgcctggtggcgttctctga	326
Db	702	cggtgcgcacgcccattctggtctctggaggtgatcggcgcgccggcgagggcgtgca	761
Qy	327	ccggacggacaagctcgtgcgctcggcgagcacacgcctgggtgacttcgacgggcaa	386
Db	762	gccagccagtacttcgggctcggcgcgggcaggctcacggtgaccgagttcgctctacgg	821
Qy	387	cctggag	393
Db	822	caaggag	828

```

RESULT 7
US-09-017-706-7
; Sequence 7, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTIDASE, VECTOR CONTAINING SAID GENE AND
; TITLE OF INVENTION: TRANSFORMANT
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO. 7

```

	Query Match	12.28;	Score 53.4;	DB 3;	Length 1981;
	Best Local Similarity	51.04;	Pred. No. 0.011;		
	Matches 126; Conservative	0;	Mismatches 121;	Indels	Gaps
Qy	147	gacctgggcgttcggccgcggagctgcttggaacgggtcacaaccgcagcgcgacgg	206		
Dd	582	gatctccgactacctggtcgaccctggtccagatggcgctcaagggtctgcggtcgatgc	641		
Qy	207	cgatgtgcacatcgccgcagccagccagggcctcgagaatgtcccacatccggtcca	266		
Dd	642	ggccaagcacatcagccgcagcagctgggtgccatcatcgacagcgtcaacgcgcgcac	701		
Qy	267	ggtcggcgcgacacctgctgcgtgttctccggcggggagcgcgcacgcctgggtggcgttccctcga	326		
Dd	702	cvgttccgcacggccattctggttctctgagtgtagtcggcgcgcggcggcggagcggtgca	761		
Qy	327	ccggagcagacaagctcgtgcgcctcggccgagagcacacgcctgggtgacttctgcagcgcaa	386		
dD	762	gcccaagcagtacttcgggctcggcggcgaggtcagcgtgacccgagtctgcctaagg	821		

```

Qy      387 cctggag 393
      | | | | |
Db      822 caaggag 828

RESULT      8
US-09-017-706-8
; Sequence 8, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPENTAOSE, VECTOR CONTAINING SAID GENE AND
; TITLE OF INVENTION: TRANSFORMANT
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID: POS3410L139
US-09-017-706-8

```

RESULT 9  
US-08-034-650-9  
: Sequence 9, Application US/08034650  
: Patent No. 5641671  
: GENERAL INFORMATION:  
: APPLICANT: BOS, Jannetje W.  
: APPLICANT: FRENKEN, Leon G.  
: APPLICANT: VERRIPS, Cornelis T.  
: APPLICANT: VISSER, Christiaan  
: TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE

;; TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS

;; NUMBER OF SEQUENCES: 13

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN

;; STREET: 1615 L. Street, N.W.

;; CITY: Washington

;; STATE: D.C.

;; COUNTRY: USA

;; ZIP: 20036-5601

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/034,650

;; FILING DATE:

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/727,235

;; FILING DATE: 03-JUL-1991

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Kokulis, Paul N.

;; REGISTRATION NUMBER: 16,773

;; REFERENCE/DOCKET NUMBER: PNK/5970/91731

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (202) 861-3000

;; TELEFAX: (202) 822-0944

;; TELEX: 6714627 CUSH

;; INFORMATION FOR SEQ ID NO: 9:

;; SEQUENCE CHARACTERISTICS:

;; TYPE: nucleic acid

;; LENGTH: 2900 base pairs

;; STRANDEDNESS: double

;; TOPOLOGY: linear

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: 483..1556

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: 1559..2617

;; US-08-034-650-9

Query Match 11.3%; Score 49.6; DB 1; Length 2900;

Best Local Similarity 50.9%; Pred. No. 0.06;

Matches 118; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 142 cctgtgacctggcggttcggccgagctgtgtgacgggctcaacagcccgagcgcc 201

DB 2081 CCGGCCGCATCGAGTCGCGCTCGATCAGCGCGCGCGCTGGCCGACCGCGCTCGGC 2140

QY 202 gacggcgatgtgcacatcgcccgaccgagcccgagcggtcggtcgagatgtccacatccgg 261

DB 2141 GAGTGGCGGAGCGGCTTCTTCGGCGACGACGACGCGCGCGCGCGCATGACCTCGAACGG 2200

QY 262 ctccaggtcgccgacgctgtgttcgggggggacggacccgctgtgtggttc 321

DB 2201 ATCCGGATCGCAACGACACACACCGCTGAGCCTGAGCAGAAAGCGCGCGGTTCGCGG 2260

QY 322 ctccagcgacggacaagctgtgcctcgctcgccagggagcacacgctgggtg 373

DB 2261 CTCGAGCGCAGCTGACGCGCGGACGAGCGCGCGACGAGCGCGCTGCTGATG 2312

RESULT 10

US-08-449-015-9

; Sequence 9, Application US/08449015

; Patent No. 5804409

; GENERAL INFORMATION:

; APPLICANT: BOS, Jannetje W.

; APPLICANT: FRENKEN, Leon G.

; APPLICANT: VERRIPS, Cornelis T.

;; APPLICANT: VISSER, Christiaan

;; TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE

;; TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS

;; NUMBER OF SEQUENCES: 13

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN

;; STREET: 1615 L. Street, N.W.

;; CITY: Washington

;; STATE: D.C.

;; COUNTRY: USA

;; ZIP: 20036-5601

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/449,015

;; FILING DATE: 24-MAY-1995

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/727,235

;; FILING DATE: 03-JUL-1991

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Kokulis, Paul N.

;; REGISTRATION NUMBER: 16,773

;; REFERENCE/DOCKET NUMBER: PNK/5970/91731

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (202) 861-3000

;; TELEFAX: (202) 822-0944

;; TELEX: 6714627 CUSH

;; INFORMATION FOR SEQ ID NO: 9:

;; SEQUENCE CHARACTERISTICS:

;; TYPE: nucleic acid

;; LENGTH: 2900 base pairs

;; STRANDEDNESS: double

;; TOPOLOGY: linear

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: 483..1556

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: 1559..2617

;; US-08-449-015-9

Query Match 11.3%; Score 49.6; DB 1; Length 2900;

Best Local Similarity 50.9%; Pred. No. 0.06;

Matches 118; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 142 cctgtgacctggcggttcggccgagctgtgtgacgggctcaacagcccgagcgcc 201

DB 2081 CCGGCCGCATCGAGTCGCGCTCGATCAGCGCGCGCGCTGGCCGACCGCGCTCGGC 2140

QY 202 gacggcgatgtgcacatcgcccgaccgagcccgagcggtcggtcgagatgtccacatccgg 261

DB 2141 GAGTGGCGGACCGCTTCTTCGGCGACGACGACGCGCGCGCGCATGACCTCGAACGG 2200

QY 262 ctccaggtcgccgacgctgtgttcgggggggacggacccgctgtgtggttc 321

DB 2201 ATCCGGATCGCAACGACACACCGCTGAGCCTGAGCAGAAAGCGCGCGGTTCGCGG 2260

QY 322 ctccagcgacggacaagctgtgcctcgctcgccagggagcacacgctgggtg 373

DB 2261 CTCGAGCGCAGCTGACGCGCGGACGAGCGCGCGCGCTGCTGATG 2312

RESULT 11

US-08-387-942C-7

; Sequence 7, Application US/08387942C

; Patent No. 5939289

; GENERAL INFORMATION:

; APPLICANT: ERTESVAG, HELGA





TITLE OF INVENTION: Specific Polyketides  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steven F. Weinstein  
STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
STREET: Park Rd  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/439,009A  
FILING DATE: 11-May-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Casuto, Dianne  
REGISTRATION NUMBER: 40,943  
REFERENCE/DOCKET NUMBER: 4952.US.D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847-938-3137  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea  
STRAIN: NRRL 2338

Qy	90	ggtcgcgatcggatgcatcgatgaacgttccacettcccgcgatgccctgtgac	149
Db	6143	GGCCACCGCGCCTTCGGCCATCCGCTCCGCCACATCCCTCCACGCCACGGCGT-CG	6085
Qy	150	ctggcggttcggcgcgagctgctgagcgggtcaacaggccgagcggcgacgcga	209
Db	6084	CGGGAGTCCGTCTCTCCCGCTCTCTGGCGAGCCCTCAGAGTAGGCGTTGCCGGCG	6025
Qy	210	tgtgcacatcgggccgcgacgagccgagggcctcgagatgtccacatccgggtccaagt	269
Db	6024	CGTAGCGCGAGCCCGGGGCGCGAAGAGCGCGAGGCGAAGACGACACGAAG	5965
Qy	270	cggcgcgaccgtgcgtgcttccggcgggagcgcaccgctggtgcgttctccaccg	329
Db	5964	CGGTCAAGTCCAGCTCGCGCTCAGCTCGTGCAGTTCCGCCGCCGAGACACCTTGGCGC	5905
Qy	330	gacggacaagctgtgcgctcggccaggagcacacgctgggtgacttcgacg	382
Db	5904	GACTTCCCGCTCGATGGCTCCGCGGTGAGGGTGTCCACGGTCCGCTCGTCG	5852

US-08-439-009A-1/c  
; Sequence 1, Application US/08439009A  
; Patent No. 6004787  
; GENERAL INFORMATION:  
; APPLICANT: Donadio, S  
; APPLICANT: Katz, L  
; APPLICANT: McAlpine, J B  
; TITLE OF INVENTION: Method of Di

```

OTHER INFORMATION: /ade1= FUNCTION
FEATURE:
NAME/KEY: CDS
LOCATION: 744..11219
OTHER INFORMATION: /function= "gene= "eryA""
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
OTHER INFORMATION: 6-deoxyerythronolide B""
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..1868
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1998..2198
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2250..3626
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3831..4811
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase 2 domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5574..6125
OTHER INFORMATION: /function= "approximate span of

```



Search completed: July 18, 2002, 11:24:25  
Job time: 10855 sec

[illegible]

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 07:26:40 ; Search time 4821.96 Seconds  
(without alignments)  
1225.988 Million cell updates/sec

Title: US-09-749-185-1

Perfect score: 438

Sequence: 1 atgcgcgagtcggttcaagc.....aggagcagaacccgcgtga 438

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	58.8	13.4	925	12	CNS0091P
C 2	57.8	13.2	925	12	CNS0091P
C 3	53	12.1	928	12	AQ856496
C 4	53	12.1	935	12	CNS006XX
C 5	52.6	12.0	556	12	AQ851215
C 6	52.2	11.9	414	12	AQ847989
C 7	51.2	11.7	555	12	AQ851019
C 8	50.6	11.6	474	12	AQ851336
C 9	50	11.4	755	12	CNS02DOV
C 10	49.8	11.4	558	10	C97336
C 11	49.6	11.3	442	12	AQ849196
C 12	49.6	11.3	1009	12	CNS010EW
C 13	49.6	11.3	1129	12	CNS02AGC
C 14	49.4	11.3	629	12	BH017892
C 15	49.4	11.3	1313	12	AG032885
C 16	49.2	11.2	912	10	BE418762
C 17	49	11.2	677	12	BH019193

18	48.8	11.1	437	10	BE400231
19	48.8	11.1	440	12	AQ852321
C 20	48.2	11.0	1101	12	CNS017SY
C 21	48	11.0	390	12	AQ846145
C 22	48	11.0	732	12	BH018404
C 23	47.8	10.9	380	12	AQ846152
C 24	47.8	10.9	578	12	BH020269
C 25	47.8	10.9	705	12	BH019836
C 26	47.6	10.9	395	12	AQ852043
C 27	47.6	10.9	631	9	BH635178
C 28	47.6	10.9	723	10	BH730817
C 29	47.4	10.8	609	12	BH019916
C 30	47.4	10.8	1339	10	BH458211
C 31	47.2	10.8	615	10	BH529148
C 32	47.2	10.8	680	10	BH859785
C 33	47.2	10.8	1521	10	BH844055
C 34	47	10.7	992	12	AG057280
C 35	46.8	10.7	452	9	AL588142
C 36	46.8	10.7	818	9	AJ395670
C 37	46.8	10.7	839	12	CNS004NB
C 38	46.8	10.7	932	12	CNS0072Q
C 39	46.8	10.7	1625	12	AG043477
C 40	46.6	10.6	631	9	AW911065
C 41	46.6	10.6	828	9	BE040853
C 42	46.6	10.6	925	12	CNS04EIN
C 43	46.6	10.6	935	12	CNS006XK
C 44	46.6	10.6	1085	12	AG152796
C 45	46.6	10.6	1296	12	AG032939

## ALIGNMENTS

RESULT 1  
CNS0091P/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

AL053013

AL053013.1 GI:4934461

GSS.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;

BP 191 91006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

- Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

PI and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1..925

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="RPCI-98"













[illegible]



```

Db 487 GGTCCCGTCTCACCAGACACCACTCGGTGGGGCGAGCGCGCTCCGTCCGGGTGGCGAGA 546
Qy 304 gcaccgtgtgtggttcctcgcagcagcgagcaagctcgtgcgcgtcgcgcagagagcac 363
Db 547 ACCGACAGTCCGGCTCGGGCCCTCGCTGCTAGCGCGCCGCGCAAGGGCGACCGGGCC 606
Qy 364 acgctgggtgacttcgacgcaacctgaggagacga 399
Db 607 CGATTGGTGCAGCGCGCTCGGCTCTGACACATCGCA 642

RESULT 14
BH017892
LOCUS BH017892 629 bp DNA linear GSS 25-MAY-2001
DEFINITION L11211.d.Hygt3.1 Leishmania major Friedlin Cosmid Genomic Library
Leishmania major genomic clone L11211, DNA sequence.
ACCESSION BH017892
VERSION BH017892.1 GI:14196598
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 629)
AUTHORS Myler,P.J., Vogt,C., Cawthra,J., Klacking,M., Marty,A., Mack,J.,
Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G., Aggarwal
,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.
TITLE Leishmania major Friedlin Cosmid End Sequences
JOURNAL Unpublished (2000)
COMMENT Contact: Myler PJ
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerpj@sbri.org
Seq primer: Hygt3
Class: cosmid ends.
FEATURES
source
1. .629
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="L11211"
/clone_lib="Leishmania major Friedlin Cosmid Genomic
Library"
/lab_host="E. coli ED8767"
/note="Vector: cLHYG; Site_1: BamHI; Genomic DNA from
Leishmania major Friedlin was partially digested with
Sau3AI, size selected, and ligated with BamHI-digested
cLHYG cosmid vector DNA. 9216 clones were picked and
arrayed. Library construction is described in Ivens et
al., Genomics Research, 8:135-145 (1998). The cLHYG
vector (Acc.No. CVU59231) is described in Ryan et al,
Gene, 131:145-150 (1993)."
BASE COUNT 140 a 189 c 229 g 67 t 4 others
ORIGIN

Query Match 11.3%; Score 49.4; DB 12; Length 629;
Best Local Similarity 47.1%; Pred. No. 20;
Matches 146; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

Qy 7 gagtcggttcaagcagaggtcatgatgagcttctcgtctccagagagagctcgttcgt 66
Db 297 GACGCCATGGCGGGAGCTGCTGCTCTGGACGGCGCACACGACGAGGACGGCGACGGC 356
Qy 67 attccggtgagctccgatacagaggtcgcgatccgtatgccatcccgatgacgttccac 126
Db 357 GCGAGAGTGTGCTGTGGACGGCGCAGACGAGGACGCCACCGCGCGGAGCTGCTG 416
Qy 127 cttccggcgatgcccctgtgacctggggttccggccgagctgctgctggagcggtc 186
Db 417 CTGTGGACGGCGCANACGAGGACGCGGACGCGCGCGGAGCTGCTGCTGGACGGCGCA 476

```

```

Qy 187 aacagcccgagcggcgagcgtgtgcacatcgcccgacccgagcccgagggcgctcgga 246
Db 477 NAGGACGAGACGCGGCGGCGGAGCTGCTGTGGACGGCGCANAGGACGAGACGCC 536
Qy 247 gatctcacatcccggtctccaggtcggcgagcagctgctgttccggcggggagcgaca 306
Db 537 GATGGCGGAGCTGCTGCTGAGCGGCGCAGACGAGGACGCGCGCGGCGGAGCTG 596
Qy 307 ccgctggtgg 316
Db 597 CTGCTGCTGG 606

RESULT 15
AG032885/c
LOCUS AG032885 1313 bp DNA linear GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-007M05.F, genomic survey sequence.
ACCESSION AG032885
VERSION AG032885.1 GI:16559758
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-007M05.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1313)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
DIRECT SUBMISSION
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
Location/Qualifiers
SOURCE 1. .1313
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-007M05.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 65 a 389 c 502 g 132 t 225 others
ORIGIN

Query Match 11.3%; Score 49.4; DB 12; Length 1313;
Best Local Similarity 42.5%; Pred. No. 20;
Matches 158; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

Qy 47 ccgaggagctcgttcgttatccgttgagctcccgatcacgagtcgagtcgcatgtatg 106
Db 743 CCGCCCNCCNNCCNCCGNCNCCGNCNCCGNCNCCGNCNCCGNCNCCGNCNCCGNC 684
Qy 107 ccattccggatgacgttccaccttcccggcgatgcccctgtgacctgtgagcgtgcgcg 166
Db 683 GCGCNCNCACACCCCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 624

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:21:23 ; Search time 4085.76 Seconds  
(without alignments)  
2084.582 Million cell updates/sec

Title: US-09-749-185-2  
Perfect score: 407  
Sequence: 1 atgagcttcctcgtctccga.....gaggagcagacgcggctg 407

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
-----						

1	407	100.0	407	6	AX007218	AX007218 Sequence
2	407	100.0	438	6	AX007216	AX007216 Sequence
3	407	100.0	1513	1	D50051	D50051 Streptomyces
4	298.2	73.3	407	6	AX007224	AX007224 Sequence
5	298.2	73.3	438	1	AF195772	AF195772 Streptomyces
6	259.8	63.8	407	6	AX007222	AX007222 Sequence
7	259.8	63.8	408	1	AF195770	AF195770 Streptomyces
8	256.6	63.0	15441	1	SCQ11	AL096823 Streptomyces
9	239	58.7	407	6	AX007220	AX007220 Sequence
10	239	58.7	566	1	AF195771	AF195771 Streptomyces
11	110.4	27.1	35284	1	SCE19A	AL096852 Streptomyces
12	106.8	26.2	38640	1	SCL2	AL137778 Streptomyces
13	106.8	22.3	1410	1	STMLACBG	M28303 S.albus bet
14	85.2	20.9	32704	1	SC8A11	AL391041 Streptomyces
15	79.2	19.5	37245	1	SC5F2A	AL049587 Streptomyces
16	75.4	18.5	45624	6	AX089419	AX089419 Sequence
17	75.4	18.5	50000	6	AX089416	AX089416 Sequence
18	63	15.5	36583	1	SC5H1	AL049863 Streptomyces
19	58.4	14.3	40356	1	SCL6	AL159139 Streptomyces
20	58.2	14.3	110000	2	LMFLCHR32_06	Continuation (7 of
21	55.4	13.6	939	1	D87847	D87847 Streptomyces
22	55.4	13.6	37200	1	SCE63	AL035640 Streptomyces
23	55.2	13.6	36849	1	SC1C3	AL023702 Streptomyces
24	54.8	13.5	80161	1	AY007564	AY007564 Saccharop
25	54.8	13.5	80161	6	AR165018	AR165018 Sequence
26	54.8	13.5	215050	1	AL646084	AL646084 Ralstonia
27	54.4	13.4	1549	1	XANHEMA	D15073 Xanthomonas
28	54.4	13.4	1549	6	E08655	E08655 DNA encodin
29	54.4	13.4	10344	1	AE004767	AE004767 Pseudomon
30	54	13.3	103843	2	AP004006	AP004006 Oryza sat
31	54	13.3	110000	2	LMFLCHR36_07	Continuation (8 of
32	53.8	13.2	1778	1	D87846	D87846 Streptomyces
33	53.8	13.2	7235	1	SGNUSG	X72787 S.griseus n
34	53.6	13.2	122715	2	AC092558	AC092558 Oryza sat
35	53.6	13.2	438	6	AX007216	AX007216 Sequence
36	53.6	13.2	1513	1	D50051	D50051 Streptomyces
37	53.4	13.1	1981	6	AR103043	AR103043 Sequence
38	53.4	13.1	1981	6	AR103044	AR103044 Sequence
39	53.4	13.1	1981	6	AR103045	AR103045 Sequence
40	53.4	13.1	1981	6	AR103046	AR103046 Sequence
41	53.4	13.1	1981	6	AR103047	AR103047 Sequence
42	53.4	13.1	1981	6	AR103048	AR103048 Sequence
43	53.4	13.1	1981	6	E24180	E24180 Alpha-amyla
44	53.4	13.1	1981	6	E24181	E24181 Alpha-amyla
45	53.4	13.1	1981	6	E24182	E24182 Alpha-amyla

ALIGNMENTS

RESULT	1	AX007218	Sequence 3 from Patent WO00000613.	407 bp	DNA	linear	PAT 06-SEP-2000
AX007218	LOCUS	AX007218	Sequence 3 from Patent WO00000613.				
	DEFINITION	AX007218	Sequence 3 from Patent WO00000613.				
	ACCESSION	AX007218	Sequence 3 from Patent WO00000613.				
	VERSION	AX007218.1	GI:9995084				
	KEYWORDS		Streptomyces griseus.				
	SOURCE		Streptomyces griseus				
	ORGANISM		Streptomyces griseus				
	REFERENCE		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
	AUTHORS		Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.				
	TITLE		1 (bases 1 to 407)				
			Kraal,B., Luiten,R.G. and Van Wezel,G.P.				
			Reducing branching and enhancing fragmentation in culturing				
			filamentous microorganisms				
	JOURNAL		Patent: WO 000613-A 3 06-JAN-2000;				
			UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA				
			(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS				
	FEATURES		VAN (NL)				
			Location/Qualifiers				
			1..407				
			/organism="Streptomyces griseus"				
			/db_xref="taxon:1911"				

```
gene      1. .405
CDS       /gene="ssgA"
          1. .>405
          /gene="ssgA"
          /note="strain ATTC of Streptomyces griseus"
          /codon_start=1
          /transl_table=11
          /product="SsgA"
          /protein_id="CAC07385.1"
          /db_xref="GI:9995085"
          /translation="MSFLVSEELSFRIPVELRYEVDPAIRMTFHPGDAPVYNAFG
          KELLGLNSPGSDGVHIGPTEPEGLGVHIRLQVGADRALFRAGTAPLVAFLDRTD
          KLVLQGHITLGDGNLEDALEQNAQ"
BASE COUNT      60 a 139 c 142 g 66 t
ORIGIN
```

```
Query Match      100.0%; Score 407; DB 6; Length 407;
Best Local Similarity 100.0%; Pred. No. 9.9e-43;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgagcttcctcgtccgaggagctcgttcctcgattccggtgagctccgataccag 60
Db 1 ATGAGCTTCCTCGTCTCCGAGGAGCTCTCGTTCGGTATTCGGTGGAGCTCCGATACGAG 60

Qy 61 gtccgcatcgtatgcattccgagatgaacttcacattcccgccgcatgccccgtgtgacc 120
Db 61 GTCGGCATCCGATGCCATCCGATGACGTTCCACCTTCCCGCGGATGCCCTGTGACC 120

Qy 121 tggcgcttcggccgcgagctgctgtgacgggctcaacagcccgagcgagcgagcgat 180
Db 121 TGGCGCTTCGGCCGCGAGCTGCTGTGGACGGGCTCAACAGCCGACGCGCGCGAT 180

Qy 181 gtcaacatcgcccgacggagccggagggcctcgagagatgcacatccgggtccagggtc 240
Db 181 GTGCACATCGCCCGACCGAGCCGAGGGCTCGGAGATGTCACATCCGGCTCCAGGTC 240

Qy 241 ggcgcgacccgtgcgtgttcctccggggagcgacacccgctggtggcgtccctcgaccgg 300
Db 241 GGCgcGACccGTgcGtGtTccTccGggGgagcGacAcCCcgctGgtGgcGtTccTcgAcCgg 300

Qy 301 acggacaagctgtgcgctgcggccaggagcacacgctgggtgacttcgacggcgaacctg 360
Db 301 ACGGACAAGCTGCTGCGCTGCGCGTGGCGAGGACACACGCTGGGTGACTTCGACGGCAACCTG 360

Qy 361 gaggacgactggcgccatcctccgagagagacagacccggctg 407
Db 361 GAGGACGACTGGGCGCATCTCTCGCGAGGAGCAGAAACCGCGGCTG 407
```

```
RESULT 2
LOCUS      AX007216                      438 bp      DNA      linear      PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0000613.
ACCESSION  AX007216
VERSION     AX007216.1 GI:9995082
KEYWORDS   .
SOURCE      Streptomyces griseus.
ORGANISM   Streptomyces griseus
REFERENCE   1 (bases 1 to 438)
AUTHORS    Kraal,B., Luiten,R.G. and Van Wezel,G.P.
TITLE      Reducing branching and enhancing fragmentation in culturing
           filamentous microorganisms
JOURNAL    Patent: WO 0000613-A 1 06-JAN-2000;
           UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA
           (NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS
           VAN (NL)
FEATURES   Location/Qualifiers
            source      1. .438
                        /organism="Streptomyces griseus"
                        /db_xref="taxon:1911"
```

```
gene      1. .408
CDS       /gene="ssgA"
          1. .>408
          /gene="ssgA"
          /note="strain ATTC of Streptomyces griseus"
          /codon_start=1
          /transl_table=11
          /product="SsgA"
          /protein_id="CAC07384.1"
          /db_xref="GI:9995083"
          /translation="MRESVQAEVMMFLVSEELSFRIPVELRYEVDPAIRMTFHP
          GPAPVTWAFGRILLDGLNSPGSDGVHIGPTEPEGLGVHIRLQVGADRALFRAGTA
          PLVAFIDRTDKLVLPGQEHITLGDGNLEDALEGR"
BASE COUNT      68 a 145 c 153 g 72 t
ORIGIN

Query Match      100.0%; Score 407; DB 6; Length 438;
Best Local Similarity 100.0%; Pred. No. 9.7e-43;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgagcttcctcgtccgaggagctcgttcctcgattccggtgagctccgataccag 60
Db 31 ATGAGCTTCCTCGTCTCCGAGGAGCTCTCGTTCGGTATTCGGTGGAGCTCCGATACGAG 90

Qy 61 gtccgcatcgtatgcattccgagatgaacttcacattcccgccgcatgccccgtgtgacc 120
Db 91 GTCGGGATCCGATGCCATCCGATGACGTTCCACCTTCCCGCGGATGCCCTGTGACC 150

Qy 121 tggcgcttcggccgcgagctgctgtgacgggctcaacagcccgagcgagcgagcgat 180
Db 151 TGGCGCTTCGGCCGCGAGCTGCTGTGGACGGGCTCAACAGCCGACGCGCGCGAT 210

Qy 181 gtcaacatcgcccgacggagccggagggcctcgagagatgcacatccgggtccagggtc 240
Db 211 GTGCACATCGCCCGACCGAGCCGAGGGCTCGGAGATGTCACATCCGGCTCCAGGTC 270

Qy 241 ggcgcgacccgtgcgtgttcctccggggagcgacacccgctggtggcgtccctcgaccgg 300
Db 271 GGCgcGACccGTgcGtGtTccTccGggGgagcGacAcCCcgctGgtGgcGtTccTcgAcCgg 330

Qy 301 acggacaagctgtgcgctgcggccaggagcacacgctgggtgacttcgacggcgaacctg 360
Db 331 ACGGACAAGCTGCTGCGCTGCGCGTGGCGAGGACACACGCTGGGTGACTTCGACGGCAACCTG 390

Qy 361 gaggacgactggcgccatcctccgagagagacagacccggctg 407
Db 391 GAGGACGACTGGGCGCATCTCTCGCGAGGAGCAGAAACCGCGGCTG 437
```

```
RESULT 3
LOCUS      D50051                      1513 bp      DNA      linear      BCT 19-MAY-1999
DEFINITION Streptomyces griseus DNA for ssgA, complete cds.
ACCESSION  D50051
VERSION     D50051.1 GI:1772323
KEYWORDS   ssgA.
SOURCE      Streptomyces griseus (strain:B2682) DNA.
ORGANISM   Streptomyces griseus
REFERENCE   1 (sites)
AUTHORS    Kawamoto,S. and Ensign,J.C.
TITLE      Cloning and characterization of a gene involved in regulation of
           sporulation and cell division of Streptomyces griseus
JOURNAL    Actinomycetol. 9, 136-151 (1995)
REFERENCE   2 (sites)
AUTHORS    Kawamoto,S. and Ensign,J.C.
TITLE      Isolation of mutants of Streptomyces griseus that sporulate in
           nutrient rich media
JOURNAL    Actinomycetol. 9, 124-135 (1995)
REFERENCE   3 (sites)
AUTHORS    Kawamoto,S., Watanabe,H., Hesketh,A., Ensign,J.C. and Ochi,K.
```



**TITLE** Expression analysis of the *ssgA* gene product, associated with sporulation and cell division in *Streptomyces griseus*

**JOURNAL** Microbiology (Reading, England) 143 (Pt 4), 1077-1086 (1997)

**MEDLINE** 97286526

**PUBMED** 9141673

**REFERENCE** 4 (bases 1 to 1513)

**AUTHORS** Shinichi, K. and Ensign, J.

**TITLE** Cloning and characterization of a gene involved in sporulation and cell division of *Streptomyces griseus*

**JOURNAL** Unpublished (1995)

**REFERENCE** 5 (bases 1 to 1513)

**AUTHORS** Shinichi, K.

**TITLE** Direct Submission

**JOURNAL** Submitted (06-APR-1995) Kawamoto Shinichi, National Food Research Institute, Biomolecular Transfection Laboratory, Kannondai 2-1-2, Tsukuba, Ibaraki 305, Japan (Tel:0298-38-8124, Fax:0298-38-7996)

**FEATURES**

source 1..1513

    /organism="Streptomyces griseus"

    /strain="B2682"

    /db\_xref="taxon:1911"

    385..389

    392..802

    /citation={4}

    /codon\_start=1

    /transl\_table=11

    /product="ssgA"

    /protein\_id="BAA21558.1"

    /db\_xref="GI:2281004"

    /translation="MMSFLVSEELSPRIPELVRYEVDPIAYRMTHFLPGDAPVTWAF  
GRELLDGLNSPGDGVHIGTEPEGLGVHIRLOVGADRALFRAGTAPLVAFLDRD  
DKLVLPGQHTLGFDFDGNLEDALGRILAEQNAG"

    stem\_loop 839..875

    stem\_loop 863..968

BASE COUNT 251 a 539 c 493 g 230 t

ORIGIN

Query Match 100.0%; Score 407; DB 1; Length 1513;  
Best Local Similarity 100.0%; Pred. No. 7e-43;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgagttctctcgtccgaggagctctcgttccgtattccggtggagctccgatacag 60  
|||||  
Db 395 ATGAGCTTCCTCGTCCGAGGAGCTCTCGTTCGTATTCCGGTGGAGCTCCGATACGAG 454  
|||||

Qy 61 gtccggatccgtatgccatccggatgacgttccacattcccgcgatgccctgtgacc 120  
|||||  
Db 455 GTCCGCATCCGTATGCCATCCGATGACGTTCCACCTTCCGCGCATGCCCTGTGACC 514  
|||||

Qy 121 tggcgcttcggccgagctgctgtgacgggtcaacagccgagcgacgacgacgac 180  
|||||  
Db 515 TGGCGTTCGGCCGCGAGCTGCTGTGACGGGCTCAACAGCCGACGCGGCGGCGAT 574  
|||||

Qy 181 gtccacatcgcccgacccgagggcctcgagatgtccacatccggtccaggtc 240  
|||||  
Db 575 GTGCACATCGCCCGACCGAGCCGAGGGCTCGGAGATGTCACATCCGGCTCCAGGTC 634  
|||||

Qy 241 ggcgcgacgtcgctgttccgggggacggcaaccgctggtggcttccctgacgg 300  
|||||  
Db 635 GCGCGGACCGCTGCGCTGTTCGGGGGGGACGCGACCGCTGTGGCGTTCTTCGACCG 694  
|||||

Qy 301 acggacaagctcgctgcccgtcggcagagacacacgctgggtgacttcgacggcaacctg 360  
|||||  
Db 695 ACGGACAAGCTCGTGGCGCTCGCCAGAGACACACGCTGGGTGACTTCGAGGGCAACTG 754  
|||||

Qy 361 gaggacgactggcccgcatctcctcgcgagagagagagacccggctg 407  
|||||

Db 755 GAGGACGCACTGGGCGCGCATCTCTCGCGAGGAGGAGAGACGCGGCGCTG 801  
|||||

RESULT 4  
AX007224

**LOCUS** AX007224

**DEFINITION** Sequence 9 from Patent WO0000613.

**ACCESSION** AX007224

**VERSION** AX007224.1 GI:9995090

**KEYWORDS** Streptomyces netropsis.

**SOURCE** Streptomyces netropsis.

**ORGANISM** Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 407)

**REFERENCE** Kraal, B., Luiten, R.G. and Van Wezel, G.P.

**AUTHORS** Reducing branching and enhancing fragmentation in culturing filamentous microorganisms

**TITLE** Patent: WO 0000613-A 9 06-JAN-2000;

**JOURNAL** UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA (NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS VAN (NL)

**FEATURES**

Location/Qualifiers

source 1..407

    /organism="Streptomyces netropsis"

    /db\_xref="taxon:55404"

    1..405

    /gene="ssgA"

    1..>405

    /gene="ssgA"

    /notes="strain ATCC of Streptomyces netropsis"

    /codon\_start=1

    /transl\_table=11

    /product="ssgA"

    /protein\_id="CAC07388.1"

    /db\_xref="GI:9995091"

    /translation="MSFLVSEELSPRIPELVRYEVDPIAYRMTHFLPGDAPVTWAF  
RELLDGLNRPDGVHIAPTDPEGLSDVSIQLVGADRALFRAGAPPLVAFLDRTD  
KSVPLGQEOTLGFDFEDSLAEALKILAEQNAG"

BASE COUNT 63 a 151 c 131 g 62 t

ORIGIN

Query Match 73.3%; Score 298.2; DB 6; Length 407;  
Best Local Similarity 83.3%; Pred. No. 5e-29;  
Matches 339; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 1 atgagttctctcgtccgaggagctctcgttccgtattccggtggagctccgatacag 60  
|||||  
Db 1 ATGAGCTTCCTCGTCCGAGGAGCTCTCCTCAAGATCCCACTCGAACTCGGATACGAG 60  
|||||

Qy 61 gtccggatccgtatgccatccggatgacgttccacattcccgcgatgccctgtgacc 120  
|||||  
Db 61 ACCCGGATCCCTACGCGGTGCGGATGACCTTCCACCTCCCGCGAGACGCGCTGTGACC 120  
|||||

Qy 121 tggcgcttcggccgagctgctgtgacgggtcaacagccgagcgacgacgacgac 180  
|||||  
Db 121 TGGCGTTCGGCCGCGGAGCTGCTGTCTGACGGGATCAACCCCGGAGCGGCGGCGGAC 180  
|||||

Qy 181 gtccacatcgcccgacccgagggcctcgagatgtccacatccggtccaggtc 240  
|||||  
Db 181 GTCCACATCGCCCGACCGACCCCGAGGGCTGTGCGGAGCTCTCCATCCGGCTCCAGGTG 240  
|||||

Qy 241 ggcgcgacgtcgctgttccgggggacggcaaccgctggtggcttccctgacgg 300  
|||||  
Db 241 GCGCGGACCGCGCGCTCTTCCGTGCAAGCGCGCGCGCTGGTGGCGCTTCTTCGACCG 300  
|||||

Qy 301 acggacaagctcgctgcccgtcggcagagacacacgctgggtgacttcgacggcaacctg 360  
|||||  
Db 301 ACGGACAAGCTCGTGGCGCTCGGCTCGGTCAGGAACAGACTCTGGGTGACTTCGAGGACACGCTG 360  
|||||

Qy 361 gaggacgactggcccgcatctcctcgcgagagagagagacccggctg 407  
|||||

Db 361 GAGGCGCGCTCGCGCAAGATCTCTCGCGGAGGAGGAGAGACGCGGCGCTG 407  
|||||

RESULT 5  
AF195772

LOCUS AF195772 438 bp DNA linear BCT 01-NOV-2000  
DEFINITION Streptomyces netropsis SsgA (ssgA) gene, complete cds.  
ACCESSION AF195772  
VERSION AF195772.1 GI:11066162  
KEYWORDS  
SOURCE Streptomyces netropsis.  
ORGANISM Streptomyces netropsis  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 438)  
AUTHORS van Wezel, G.P., Rousseau, C. and Kraal, B.  
TITLE Cloning and sequencing of the Streptomyces netropsis ssgA gene  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 438)  
AUTHORS van Wezel, G.P.  
TITLE Direct Submission  
JOURNAL Submitted (18-OCT-1999) Biochemistry, University of Leiden, PO Box  
9502, Leiden 2300 RA, Netherlands  
FEATURES  
source  
1..438  
/organism="Streptomyces netropsis"  
/db\_xref="taxon:55404"  
gene  
1..438  
/gene="ssgA"  
CDS  
1..438  
/gene="ssgA"  
/note="similar to Streptomyces griseus SsgA; possible cell  
division protein"  
/codon\_start=1  
/transl\_table=11  
/product="SsgA"  
/protein\_id="AAG28483.1"  
/db\_xref="GI:11066163"  
/translation="WRESVOAEVLSFLVSELSKIPVELRYETRDPYAVRMTPLP  
GDPIWAFGRRELLDGINRPSGDGVHIAPDPEGLSDVSRVQLQVADRALFRAGAP  
PIVAFDRDTSVPLGQOTLGFDEDSLEAALGLKILAEQNAG"  
BASE COUNT 70 a 158 c 142 g 68 t  
ORIGIN

Query Match 73.3%; Score 298.2; DB 1; Length 438;  
Best Local Similarity 83.3%; Pred. No. 4.9e-29;  
Matches 339; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
Qy 1 atgagcttctctccaggagctctcggttcgattccggtgagctccgatacag 60  
Db 31 ATGAGCTTCTCTCCGAGGAGCTCTCTTCAAGATCCAGTCCAGTCCGATACGAG 90  
Qy 61 gtccggatccgtatgccatccggatgaagttccacattccggcgcatgccctgtgacc 120  
Db 91 ACCCGGATCCCTACGCGGTGCGGATGACCTTCCACCTCCCGGAGACGCGCTGTGACC 150  
Qy 121 tggcggttcggcgagagctgctgtcgagcgggtctcaacagccggagcgagcgagat 180  
Db 151 TGGCGGTTTCGGCGGGAGTGTCTGCTCGACGGGATCAACCGCCGCGAGCGCGAC 210  
Qy 181 gtccacatcggcccgaccagcggcctcgagatgccacatcggtcccggtcccggtc 240  
Db 211 GTCCACATCGCCCGACCGACCGCGAGGCGCTGTCCGAGATCTCCATCGGCTCCAGGTG 270  
Qy 241 ggcgcgaacgtgctgctgttccggcggggagcgaccctggtgcttctccgaccgg 300  
Db 271 GCGCGGACGCGCCCTCTTCCGTGCGAGCGCGCCCGCGCTGGTGCCTTCTCCGACCGC 330  
Qy 301 acggacaagctgctgcgctcgccagagacacacactgggtgacttcgacggcaacctg 360  
Db 331 ACGGACAAGTCTGCTGCGGTTCAGGAACAGACTCTGGGTGACTTCGAGGACAGCCTG 390  
Qy 361 gaggacgactggcgcgactctccgcgagagacagaaacgcgcgctg 407  
Db 391 GAGCGCGCTCGGCAAGATCTTCGCCGAGGAGGACAGAACGCCGGCTG 437

RESULT 6  
LOCUS AX007222  
DEFINITION Sequence 7 from Patent WO0000613.  
ACCESSION AX007222  
VERSION AX007222.1 GI:9995088  
KEYWORDS  
SOURCE Streptomyces goldeniensis.  
ORGANISM Streptomyces goldeniensis  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 407)  
AUTHORS Kraal, B., Luiten, R.G. and Van Wezel, G.P.  
TITLE Reducing branching and enhancing fragmentation in culturing  
filamentous microorganisms  
JOURNAL Patent: WO 0000613-A 7 06-JAN-2000;  
UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GILSBERTUS MARIA  
(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS  
VAN (NL)  
FEATURES  
Location/Qualifiers  
1..407  
/organism="Streptomyces goldeniensis"  
/db\_xref="taxon:121022"  
gene  
1..405  
/gene="ssgA"  
CDS  
1..>405  
/gene="ssgA"  
/note="strain ATCC of Streptomyces goldeniensis"  
/codon\_start=1  
/transl\_table=11  
/product="SsgA"  
/protein\_id="CAC07387.1"  
/db\_xref="GI:9995089"  
/translation="MSFLVSELSKIPVELRYETCDPVAVRLTFLPGDAPVTWAFG  
RELLIDGPRPCGDGVHIAPADPTFVEVLRVQLQVSDQAMFRVGTAPLAFDLRTD  
KIVPLGQSRSLADFDLALDLAEQNAQ"  
BASE COUNT 56 a 135 c 145 g 71 t  
ORIGIN

Query Match 63.8%; Score 259.8; DB 6; Length 407;  
Best Local Similarity 77.4%; Pred. No. 3.4e-24;  
Matches 315; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
Qy 1 alagagcttctctccaggagctctcggttcgattccggtgagctccgatacag 60  
Db 1 ATGAGCTTCTCTCCGAGGAGCTCTCTTCCGTATTCCGCTGAGGAGTCCGTTACGAG 60  
Qy 61 gtccggatccgtatgccatccggatgaagttccacattccggcgcatgccctgtgacc 120  
Db 61 ACTGTGATCCCTACGCGGTGCGGCTGACCTTTCATCTGCCGAGATGCCCGGTGACC 120  
Qy 121 tggcggttcggcgagagctgctgtcgagcgggtctcaacagccggagcgagcgagcgat 180  
Db 121 TGGCGCTTCGGCGGGAGTGTCTCATCGACGGAGGTCCGCGCGCTGCGGGGACGGGGAC 180  
Qy 181 gtgcacatcggcccgaccagcggcctcgagatgccacatcggtcccggtcccggtc 240  
Db 181 GTCCACATCGCCCGCGCGACCGGAGAGCTTCGGCGAGGTCTCTGATCCGCTTCAGGTG 240  
Qy 241 ggcgcgaacgtgctgctgttccggcggggagcgaccctggtgcttctccgaccgg 300  
Db 241 GCGAGGACACAGCGCATGTTCCGGGTGCGCACGCGCGCGCTGGTGCCTTCTTGGACCGC 300  
Qy 301 acggacaagctgctgcgctcgccagagacacacactgggtgacttcgacggcaacctg 360  
Db 301 ACGGACAAGTCTGCTGCGGTTCGGGTGAGGAGCGTTCCTCTCGGACATTCGACGCCCTGCTC 360  
Qy 361 gaggacgactggcgcgactctccgcgagagacagaaacgcgcgctg 407  
Db 361 GAGCGCGCTCGGACCGCATCTTCGCCGAGGAGGACAGAACGCCGGCTG 407





Db	5595	TGGGCCCTCGGCGCGGAGCTGCCTGGTCAGCGGGGTGGCGCGCGCTGGCGCGACGGCGAC	5654
Qy	181	gtgcacatcgcccgagccgagccggcggctcggaagtgtccacatccggctccaggctc	240
Db	5555	GTGCGCATCGCGCGGTGGAGCGGAGCCGCTGCCAGGAGTGCTGATCCGACTCAGGTC	5714
Qy	241	ggcgcgaccgtgcgtgttcccgccggcgagcgacacgcgtggtgcyttcttcgcacgg	300
Db	5715	GGCAGCACGAGGCGCTGTTCCTTCCGTTCCTCGCGCGCGCCCTGGTGGCTTCTCGACGC	5774
Qy	301	acggacaagctgtgcgctcgcccgccagagagcacacgctgggtgacttcgacggcaacctg	360
Db	5775	ACCACAAAGTTGTCGCGCTGGGCGAGAGGGCGCTCGCCGACTTCGACAGCACCTC	5834
Qy	361	gagcagcactggcgcccatctctgccgagagagcagacgcgcgctg	407
Db	5835	GACGAGCCCTGGACCGCATCTCTGCGCGAGGAACAGAGCGGGCGTG	5881
RESULT	9		
AX007220			
LOCUS	AX007220	Sequence 5 from Patent WO0000613.	PAT 06-SEP-2000
DEFINITION	AX007220	407 bp DNA linear	
ACCESSION	AX007220		
VERSION	AX007220.1	GI:9995086	
KEYWORDS			
SOURCE		Streptomyces albus G.	
ORGANISM		Streptomyces albus G.	
REFERENCE		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
AUTHORS		Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
TITLE		1 (bases 1 to 407)	
JOURNAL		Kraal,B., Luiten,R.G. and Van Wezel,G.P. Reducing branching and enhancing fragmentation in culturing filamentous microorganisms Patent: WO 000613-A 5 06-JAN-2000; UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA (NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS VAN (NL)	
FEATURES		Location/Qualifiers	
source		1..407	
		/organism="Streptomyces albus G"	
gene		1..405	
		/db_xref="taxon:1962"	
CDS		/gene="ssgA"	
		1..>405	
		/gene="ssgA"	
		/note="strain ATCC of Streptomyces albus G"	
		/codon_start=1	
		/transl_table=11	
		/product="SsgA"	
		/protein_id="CAC07386.1"	
		/db_xref="GI:9995087"	
		/translation="MSFLVSEELAIFRVELRYETVDPAVRLLFHLPGDAPVTWVFGRQLLGVGLDAAGDVRCVPVGQTATREVIHTLVQSGEALFRVKAPLIPLAFLDRTDRELIGLSRAHADFDLSHLDNALNRSIAEQASG"	
BASE COUNT	57 a	150 c	139 g
ORIGIN			61 t
Query Match		58.7%	Score 239; DB 6; Length 407;
Best Local Similarity		74.2%	Pred. No. 1.4e-21;
Matches	302; Conservative	0; Mismatches 105; Indels	0; Gaps 0;
Qy	1	atgagcttcctgcgtccgaggagcctcgttcogtatccggttgagctccgatcacag	60
Db	1	ATGAGCTTCCTCGTCTCCGAGGAGCTCGCCTTCGCGCATCCGGTGGAGCTCGGTACGAG	60
Qy	61	gtcggcgatccgatgccattcccgatgacgttcacacctcccggcgatgccctgtgacc	120
Db	61	ACCTGTCATCGTAGCGGTGCGCTGACGTTCCACCTCCCCGGAGAGCGCCCCTGTCAC	120
Qy	121	tggcgcttcggccgcgagctgctgtagcgagcggtcacagccccgagcgacgacgacqat	180

```

Db      86  ACCGTCATCGTACGGGTGCGGTACGCTTCACCTCCCGGAGAGCGCCCGGTCAAC 145
Qy      121  tggcggttcgcccgcagctctactgacgggctcaacagccgagcggcagcgagat 180
Db      146  TGGTCTTCGGCGTGAACCTCTGGTTCAGGGAGTCTCTGGACGCCGCGGGGCGGCGAC 205
Qy      181  gtgcacatcgccgcagcagccgagcggcctcgagatgtccacatccggtccagctc 240
Db      206  GTCCGGGTCTGCCGGTGGGSCAGACGCCACCGAGGAGGTGCACATCACCTCCAGGTC 265
Qy      241  ggqcgacacgctgcgctgttctcggcgccgagcagccgctggtggtggttctcgacc 300
Db      266  GGCTCCGAGCAGGCGCTCTTCGCGGTGCGCAAGCGCGCTGCTGCGCTTCTCGACCGC 325
Qy      301  acgacacgctgcgctgcgctgcgacagcagcagcctggtggtgacttcgacggcaact 360
Db      326  ACCACACAGGCTTGTGCTCGGACGCGGCGGCGCACCGCGCTCTCGACAGCCACCTC 385
Qy      361  gagacgactggccgcacatctcgcgagagcagcagcagcggctg 407
Db      386  GACGAGCTCTGAACCGCAGCTCCCGGAGGAGCAGACGCCGCGCTG 432

RESULT 11
LOCUS   SCE19A 35284 bp DNA linear BCT 16-JUL-1999
DEFINITION Streptomyces coelicolor cosmid E19A.
ACCESSION AL096852
VERSION AL096852.1 GI:5531349
KEYWORDS 4-hydroxyphenylpyruvate dioxygenase; ABC transporter; amino acid
permease; asnC-family; cysteine synthase; nicotinamidase;
osmoregulation; osmotolerance; oxidoreductase; permease;
phosphoenolpyruvate-dependent; protease; regulator; ribonuclease
PH; rph; sugar phosphotransferase system; transcriptional
regulator; transposase; tRNA Leu; zinc metalloproteinase.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Streptomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 35284)
Redenbach,M., Kieser,H.M., Denepalte,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
2 (bases 1 to 35284)
Seeger,K. and Harris,D.
Unpublished
3 (bases 1 to 35284)
James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
Direct Submission
Submitted (16-JUL-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S.coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given.

```

where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid E19A lies between and overlaps cosmids E6 and E56 on the AseI-E genomic restriction fragment.

Location/Qualifiers

1..35284

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="cosmid E19A"

1..97

/gene="SCE19A.01"

<1..97

/gene="SCE19A.01"

/note="SCE19A.01, hypothetical protein, partial CDS, len: >31 aa; unknown function, similar to members of the alkyl hydroperoxide reductase C/thiol-specific antioxidant family e.g. TR:O53226 (EMBL:AL021185), bcp, Mycobacterium tuberculosis bacterioferritin comigratory protein (157 aa), fasta scores; opt: 124 z-score: 238.1 E(): 6.2e-06, 63.3% identity in 30 aa overlap"

/codon\_start=2

/transl\_table=11

/label="SCE19A.01

/product="hypothetical protein"

/protein\_id="CAB50982.1"

/db\_xref="GI:5531350"

/translation="IVVDEGKVERALYNVRATGHVAKIKDLGI"

140..226

/note="tRNA Leu anticodon TAG, Cove score 69.78"

/product="tRNA-Leu"

complement(243..845)

/gene="SCE19A.02c"

complement(243..845)

/gene="SCE19A.02c"

/note="SCE19A.02c, conserved hypothetical protein, len: 200 aa; unknown function, similar to many e.g. SW:Y030\_MVCTU (EMBL:Z73902), Mycobacterium tuberculosis hypothetical protein (204 aa), fasta scores: opt: 633 z-score: 743.6 E(): 0, 52.0% identity in 200 aa overlap. Similar to SW:YGGV\_ECOLI (EMBL:U28377), yggv, Escherichia coli hypothetical protein (197 aa) (49.5% identity in 198 aa overlap)"

/codon\_start=1

/transl\_table=11

/label="SCE19A.02c

/product="hypothetical protein"

/protein\_id="CAB50983.1"

/db\_xref="GI:5531351"

/translation="MTRLILATRNACKITELRALADAGLPHDLVGADAVPHIPDVKE TGVYFAENALLKAHLAATGTPAVADSLGCVDLVNGAGLGFSAWAGRHGGDQANL DULLAQIADIADHRGAHFACAAALPDGTERVGEQKGLTLRHAPAGTGGFGYDPI LQPEGTRTCAELTAEEKNALSHRGKAFRALVPVVRRELIG"

complement(892..1296)

/gene="SCE19A.03c"

complement(892..1296)

/gene="SCE19A.03c"

/note="SCE19A.03c, possible secreted protein, len: 134 aa; unknown function, probable CDS suggested by GC frameplot,

positional base preference and amino acid composition.  
Contains probable N-terminal signal sequence and  
appropriately positioned PS00013 Prokaryotic membrane  
lipoprotein lipid attachment site, so may be a  
lipoprotein"

/codon\_start=1  
/transl\_table=11  
/label\_SCE19A.03c  
/product="putative secreted protein"

/protein\_id="CAB50984.1"

/db\_xref="GI:5531352"

/translation="MAASRRRLRTTAVATVATIALTGLTTCDAVDKALDCVFT  
ADATSVTEIQQAVENADDPOMEESLNSDKNLDRIQDTNDTWNKAVDLDGKAV  
DNVRTSVNGDETDLSPVTDAAAGELTKVCTP"

complement(1201..1233)

/gene="SCE19A.03c"

/note="PS00013 Prokaryotic membrane lipoprotein lipid

attachment site"

complement(1454..2191)

/gene="SCE19A.04c"

/note="rph"

complement(1454..2191)

/gene="SCE19A.04c"

/note="SCE19A.04c, rph, probable ribonuclease PH, len: 245  
aa; highly similar to many e.g. SW:RNPH\_ECOLI  
(EMBL:X00781), rph, Escherichia coli ribonuclease PH (238  
aa), fasta scores: opt: 916 z-score: 1116.8 E(): 0. 60.8%  
identity in 240 aa overlap. Contains Pfam match to entry  
PF01138 RNase\_PH, 3' exoribonuclease family and PS01277

Ribonuclease PH signature"

/codon\_start=1

/transl\_table=11

/label=rph

/product="putative ribonuclease PH"

/protein\_id="CAB50985.1"

/db\_xref="GI:5531353"

/translation="MSRIDRTPQOLRPVPTIERGSKHAEGSVLVFGDTKVLQNASV  
TEGVPWRKSGEGWVTAAYAMLPRATNTRGDRESVKRIGRTHETISRLIGSLRAV  
IDYKALGNTVVLDCVDVLOADGTRTAITGAYVALADAVAAOGRKLIKANKRPLTG  
TVSAVSGIVDGTPLDLRYEEDVRADTMNVCTGDRFVEVQGTAEAFARDELN  
TLDLATAGCTELAEQLRKALDATALER"

complement(1511..2185)

/gene="SCE19A.04c"

/note="Pfam match to entry PF01138 RNase\_PH, 3'

exoribonuclease family, score 266.70, E-value 3.1e-76"

complement(1805..1843)

/gene="SCE19A.04c"

/note="PS01277 Ribonuclease PH signature"

complement(2268..2501)

/gene="SCE19A.05c"

complement(2268..2501)

/gene="SCE19A.05c"

/note="SCE19A.05c, hypothetical protein, len: 77 aa;  
unknown function, similar to parts of many sugar permeases  
e.g. TR:P96159 (EMBL:U65013), malX, Vibrio furnissii PTS  
(phosphoenolpyruvate-dependent sugar phosphotransferase  
system) permease for glucose (523 aa), fasta scores: opt:  
217 z-score: 287.9 E(): 1.1e-08, 48.1% identity in 77 aa  
overlap. Contains Pfam match to entry PF00367 PTS\_EIIB,  
phosphotransferase system, EIIB"

/codon\_start=1

/transl\_table=11

/label\_SCE19A.05c

/product="hypothetical protein"

/protein\_id="CAB50986.1"

/db\_xref="GI:5531354"

/translation="MKAERKIVAGLGIDNIDEIGCTITRLRTEVNDPALVNEAALK  
AAGAGVVKMGTAQVIGTADTAABIEDMM"

complement(2385..2489)

/gene="SCE19A.05c"

/note="Pfam match to entry PF00367 PTS\_EIIB,  
phosphotransferase system, EIIB, score 43.00, E-value  
8.1e-11"

gene

2660..3955

/gene="SCE19A.06"

2660..3955

/gene="SCE19A.06"

/note="SCE19A.06, possible PTS transmembrane component,

len: 431 aa; similar to many PTS

(phosphoenolpyruvate-dependent sugar phosphotransferase

Query Match

Best Local Similarity 27.1%; Score 110.4; DB 1; Length 35284;

Matches 210; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 2 tgaattctctctccgagagagctctcttcgtattccgttggtgagctccgatacagag 61

Db 20142 TCAGGCTGCTCTGTGCGCCGAGAGCGGCATTCGGGTGCGCGCGCTCGGTACACCA 20201

QY 62 tcggcgatccgtatgccatccggatgacgttccacattcccggtggtccctgtgaacct 121

Db 20202 CGGACGATCCCTACGCCGTGCACATCACCTTCACATCGACTCCCGCCACCGGTGCAC 20261

QY 122 gggcttcggcgagctgctgctggagcggtctcaacagccgagcgagcgcgatg 181

Db 20262 GGACGTTCCCGCGGACCTTCGTGTGGAGGGCGCTCTCCGGCGCTCCGGCACCGGGAGC 20321

QY 182 tgacatcgcccgacggagcgagggctcgagatgtccacatccggctccaggtcg 241

Db 20322 TCGGGGTGTGGCGTCGAAGACGGAGGCGCGAGCGTCTACTCTGCGCTGAGCAGCC 20381

QY 242 ggcggagcgtgcgtgttccgggggggacggcaccgctggtggcgttctctgacccgga 301

Db 20382 CGGACGGCGACGCCCTCTCTGGAGGCGCCACGCCAGGTGTGCGCTGTGTGAGCGGA 20441

QY 302 cggacaagctgctgcgtcgccagagcagcacacgctggtaacttcgagcgcaacctg 361

Db 20442 CCTCGGGCGGTGCGCGCGGACCGAGGGCGCGAGTCTGGGATCGAGACGGGCTGG 20501

QY 362 aggaagcactggggcg 377

Db 20502 CCGAGCTGCTGCCAG 20517

misc\_feature

misc\_feature

gene

CDS

misc\_feature



David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: <http://www.sanger.ac.uk/Projects/S-coelicolor/>)  
CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>.  
CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid L2, Location/Qualifiers  
1..38640  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid L2"  
1..989  
/gene="SCL2.01"  
1..989  
/partial  
/gene="SCL2.01"  
/note="SCL2.01, hypothetical protein, len: >328 aa; similar to TR:O3236 (EMBL:Z98209) Mycobacterium tuberculosis hypothetical 49.8 kd protein MTCv174.11, len: 450 aa; fasta scores: opt: 843 z-score: 874.7 E(): 0; 45.9% identity in 296 aa overlap. Contains possible coiled-coils region"  
/codon\_start=3  
/transl\_table=11  
/product="hypothetical protein SCL2.01"  
/protein\_id="CAB70913.1"  
/db\_xref="GI:6822207"  
/translation="SWAGSPPEALAVFERKYVGLVVEITGLLEKRVKTTDLAKDAQT AVPHLEQVDAHVAVDLEALRALDQVLAVALVTRREERKAQPAKQDARGAKGKALY AEAELARDSDQWRAAGERLSLVDTWKGLPRLDKSDDELWHRFSHARSAFRRKQKH LMQDQAREARKIRKLVSEALNSNDWGTAAARYRDLMSKAAAGRAQREHDD LWRFGADQVFFFAKRSVSFAEDQASENLKLKELVTEAEKLVPTDLKSARAAFR SVNRWEATGCHVPRDARPKSRGGTSRSGPSRRPRPSGAGTTPRHARVPA"  
1010..11150  
/gene="SCL2.02"  
1010..11150  
/gene="SCL2.02"  
/note="SCL2.02, unknown, len: 46 aa"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SCL2.02"  
/protein\_id="CAB70914.1"  
/db\_xref="GI:6822208"  
/translation="MDKLSQVEQARAQNDKADKLARELEGRQALLDQALKGLHEF

misc\_feature  
gene  
CDS  
misc\_feature  
misc\_feature  
misc\_feature  
RBS  
misc\_feature  
gene  
CDS

GG"  
/note="SCL2.03c, relA, GTP pyrophosphokinase, len: 847 aa; identical to previously sequenced SW:RELA\_STRCO (EMBL:X87267) Streptomyces coelicolor GTP pyrophosphokinase (EC 2.7.6.5) RelA, 847 aa and highly similar to SW:RELA\_STRAT (EMBL:AF072829) Streptomyces antibioticus RelA protein, 841 aa; fasta scores: opt: 4969 z-score: 4906.2 E(): 0; 90.3% identity in 848 aa overlap and to SW:RELA\_CORGL (EMBL:AF038651) Corynebacterium glutamicum GTP pyrophosphokinase (EC 2.7.6.5) RelA, 760 aa; fasta scores: opt: 1613 z-score: 1594.3 E(): 0; 58.7% identity in 774 aa overlap. Contains Pfam match to entry PF01842 ACT, ACT domain and match to Prosite entry PS00017 ATP/GTP-binding site motif A (P-loop)"  
/codon\_start=1  
/transl\_table=11  
/product="GTP pyrophosphokinase"  
/protein\_id="CAB70915.1"  
/db\_xref="GI:6822209"  
/translation="MPDEAQPLTAAPKSPASASAAKPAAPQAKNDTHGPIQHAPAA PVLPKPAEQPRKPLPAERPNQAVVPAPAGAPARSSSNRVRLARLARGVORANPVN PVEPLLRVRGNDPKLETSTLRQIERAYQVAERWHRGQKRKSGDPIYTHPLAVTTIL AEIUGMPATLMAGLHDHDTEDTEYGLDRLRRDFGVDVTLVGVTKLDKVKFGAAQA ETVRKVMAMADKPRVLVILKADLRNMRMTYLRKKEQKARETLEIYAPLAHRLG MTKIWELEDLAFILPKMYDEIVRLVAERAPKRDEYLAIVTDEVQDRLRAARIKAT VTGRPKHYSYVKQMIVRGDFAEIYDLVIRVLTVDTRDCAALGTVHARWNPVPCR FKDYIAMPKNYQSLHTTVIGPGKPVLEQIHTFDMHRAEYVIAAHMKYKQEAAGV ASKYRTDAPKSGSKDDHLNDMAHLRQLLDQKETEDEGFLESLRFDLSRNEVYVF TPQGVIALPAGATPVDFAYAVHTEVHRTIGARVNGRLVPLSTLDGDLVVEVFTSK AAGAPSRDLGFCVSPARNKIRAFWFSKERDEAEQGDIAIVRAMKONLPQIRIL TGDSLVLAHEMRYSDISALYAAIGEGHSVAPNIVQKLAQALGEEAAATEIDESVPP SGRGRKRRANADPGVVVGVYDVKVLAARCTPVPDPIGIVGTGSGVSVHRSDCV NVDSLREPERITILEVEMAPTQSSVFLVATQVALDRSLLSQVTRVLSQDQVHILSAA VQTSRDRVATSRFTFEMGDPKHLGHVKAARGVEGYDYVRYVTSARRPS"  
/gene="relA"  
/note="Pfam match to entry PF01842 ACT, ACT domain, score 63.10, E-value 6e-15"  
/complement(2378..2401)  
/gene="relA"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
/complement(3273..7460)  
/note="previously sequenced region SW:SCSECAPT EMBL:X85969 S.coelicolor secD, secE & apt genes"  
/complement(3778..3781)  
/complement(3904..4568)  
/note="previously sequenced region SW:SCAPTRELA EMBL:X87267 S.coelicolor apt & relA genes"  
/gene="apt"  
/complement(3956..4453)  
/gene="apt"  
/note="SCL2.04c, apt, adenine phosphoribosyltransferase, len: 182 aa; identical to previously sequenced SW:APT\_STRCO (EMBL:X87267) Streptomyces coelicolor adenine phosphoribosyltransferase (EC 2.4.2.7) Apt, 182 aa and highly similar to SW:APT\_ECOLI (EMBL:M14040) Escherichia coli adenine phosphoribosyltransferase (EC 2.4.2.7) Apt, 183 aa; fasta scores: opt: 600 z-score: 682.8 E(): 1.2e-30; 54.8% identity in 168 aa overlap. Contains Pfam match to entry PF00156 Ribosyltran, Phosphoribosyl transferase domain and match to Prosite entry PS00103 Purine/pyrimidine phosphoribosyl transferases signature"  
/codon\_start=1  
/transl\_table=11  
/product="adenine phosphoribosyltransferase"





**AUTHORS** Redenbach,M., Kieser,H.M., Denapaiter,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.  
**TITLE** A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
**JOURNAL** Mol. Microbiol. 21 (1), 77-96 (1996)  
**MEDLINE** 97000351  
**REFERENCE** 2 (bases 1 to 32704)  
**AUTHORS** Saunders,D.C. and Harris,D.  
**JOURNAL** Unpublished  
**REFERENCE** 3 (bases 1 to 32704)  
**AUTHORS** Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (03-AUG-2000) Streptomyces coelicolor sequencing project, by the BBSRC and Beowulf Genomics  
**REFERENCE** Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: wellcome@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

**COMMENT**  
 Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web  
 (URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
 CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
 Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>.  
 CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
 Cosmid 8A11.

**FEATURES**  
 source  
 1. .32704  
 /organism="Streptomyces coelicolor A3(2)"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 /clone="cosmid 8A11"  
 1. .100  
 /gene="SC8A11.01"  
 /note="nominal overlap with Streptomyces coelicolor cosmid St9A4"  
 1. .2254  
 /gene="SC8A11.01"  
 /note="SC8A11.01"  
 /gene="SC8A11.01"  
 /note="SC8A11.01, possible transcriptional regulator (fragment), len: >750 aa; similar to TR:Q9S124 (EMBL:AL109747) Streptomyces coelicolor probable transcriptional regulator SCJ21.13, 919 aa; fasta scores: opt: 818 z-score: 913.7 E(): 0; 32.6% identity in 775 aa overlap. Contains pfam match to entry PF00196 GerE, Bacterial regulatory proteins, luxR family and match to Prosite entry P500622 Bacterial regulatory proteins, luxR family signature. Also contains a possible helix-turn-helix motif at residues 699..720 (+3.25 SD)"

**misc\_feature**  
 1. .100  
 /product="putative transcriptional regulator (fragment)"  
 /protein\_id="CAC01573.1"  
 /db\_xref="GI:9716212"  
 /translation="IGAARTSPGTYTILGTHAEVLPLEPLSPSAAEELLDADAPGLADA VRARIQRAAGNPALVELPRAAGCISPPDLDPLOTORLETAFASTDSLTRECRFL LVLAAPTAPLNQLLDVASRLAGSEVTVYALQEAVDAGLVLTGRTPERHPLMRSAL YATVADRLSTHRALETGSGPRLVHLAAATLGPDDELAGOLERPADDAQRGQ LAAVAPLRQAGELVHDPRQTGLVRAAEALASEINDRVQAQLILNLRADLAEPGRT RMLVSDKAAEPDEPQRIQMDIAAGAFDVGSTVAENLLWRAARCFQGDHL RVRQAAELEDKWKPDPAHVLTVRAYTEPYRGCTDLIARLEKLRWAGDGRGLLYL GSGSWAIGDVGATRYLAQASWRSQRLGLLARSAGSWPRLYLGOLAOARESAAE GIATALEGTEWIVLGLKATSALTAVLGEREAAARSVRELRAHSLFVPMPTASVNAQ QVEGALLFDSRAVEAYDALARFDTDPHYHSTSRWLLVDPDLVDAAAAAGNEQARE LIVPELADRLPSEMIVARTYSTAVLAPDDTAEDCYDSALSALPTWPLARLHL QHGRLLRQRNVNDAKPLRLARDEFDRVGAQPMADMAREQLRAGAGESGRRRPSKGE SLTVQERQIAELASQGLSNREIGQRLFIHSRTVGAHYRIYPRLGITSRGLSALAAL LGDDQPTSGG"  
 2042..2239  
 /gene="SC8A11.01"  
 /note="Pfam match to entry PF00196 GerE, Bacterial regulatory proteins, luxR family, score 49.10, E-value 9.7e-11"  
 2093..2176  
 /gene="SC8A11.01"  
 /note="PS00622 Bacterial regulatory proteins, luxR family signature"  
 complement(2184..2453)  
 /gene="SC8A11.02c"  
 complement(2184..2453)  
 /gene="SC8A11.02c"  
 /note="SC8A11.02c, unknown, len: 89 aa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SC8A11.02c"  
 /protein\_id="CAC01574.1"  
 /db\_xref="GI:9716213"  
 2787..3257  
 /gene="SC8A11.03"  
 2787..3257  
 /gene="SC8A11.03"  
 /note="SC8A11.03, possible regulator, len: 156 aa; similar to TR:Q9S2F7 (EMBL:AL096852) Streptomyces coelicolor putative regulator SCE19A.24, 142 aa; fasta scores: Opt: 272 z-score: 353.0 E(): 3.4e-12; 36.9% identity in 130 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative regulator"  
 /protein\_id="CAC01575.1"  
 /db\_xref="GI:9716214"  
 /translation="MSGDHHGVQAOHSAQAALLPLSLCSOMTGALEWEDVPAEFRRYD PDHLLVTRFAPGAPPTVHWGRDLLHEGLRTSGLDGVQVWADPTPDRETAWLQV NARHQAIFSLPVELEWIDRTYLHPAGTSSRLGTDAFLSKLDFEPEASSR"  
 complement(3297..7058)  
 /gene="SC8A11.04c"  
 complement(3297..7058)  
 /gene="SC8A11.04c"  
 /note="SC8A11.04c, probable secreted peptidase, len: 1253 aa; similar to TR:Q9RL54 (EMBL:AL121596) Streptomyces coelicolor probable secreted peptidase SCF51A.10, 1245 aa; fasta scores: opt: 1827 z-score: 1848.3 E(): 0; 37.2% identity in 1300 aa overlap, to TR:P95684 (EMBL:D83672) Streptomyces albobrosculus subtilisin-like protease, 1102 aa; fasta scores: opt: 1365 z-score: 1380.6 E(): 0; 33.4% identity in 1268 aa overlap and to Streptomyces coelicolor probable secreted peptidase SC8A11.16c, 1239 aa; fasta scores: opt: 1835 z-score: 1467.4 E(): 0; 46.9% identity in 1279 aa overlap. Contains pfam matches to entries PF00082 Peptidase\_S8, Subtilase family and PF02225 PA, PA

domain and matches to Prosite entries PS00136 Serine proteases, subtilase family, aspartic acid active site, PS00137 Serine proteases, subtilase family, histidine active site and PS00138 Serine proteases, subtilase family, serine active site. Also contains possible N-terminal region signal peptide sequence"

```
/codon_start=1
/transl_table=1
/product="putative secreted peptidase"
/protein_id="CAC01576.1"
/db_xref="GI:9716215"
/translation="MPIPWRSGRGLTVGIAALLVLTATAPASALPGTVPPAPTAPA
GPTVRLITDRVTVTGEGDAEVLVTPDHRSGGAHVMTGSDTVYVPDAAVPL
GSGALDERLNVTELEDGYDDARDELPLIYVTDTAARSLGARTPEGARTRALS
IRGAASIAEHSRAADFWSLTGTGDAAGGSAARSATSGGRLAGIAKWLMDGVRAT
LSDTAIOAGPDVWVGNGTGEVGVAVLDTGVDAHPDFAGRIAATASFVPDQDVTDR
NGHCTHVASFVAGTCAAGSGVEKGVAPGASLHIGKVLNDSGGSDSVLAGMVAVRD
QKAIKSMGSDSTGDTDLSEAVNWLSEAGTALFVVAAGNSPEAYITVGTAAAADA
ALTGVANGPKGVQDLADFSRGPVRVDNAVKPDLTAPGVGLAARSRYAPEGEAY
QSLSGTSMATPHVAGAAALLAAEHPDWTGQRLKALVGTGTQRFSPFDAAGSRVDV
AAAVRSTLLASGDFAOAHVPTPCQTVRRDVTYNSGPAPVALDLALSPAELEPGLF
TLSEAQVTPAHGTASVGVITHLDAEDNGAYATRLVASGADGAVLARTPVGVNKEGR
RATIALAKDHHDRPLSGTVILKDVERTAPKYVSDASGRDLRLSPSYVWMSA
VPGVDGTHTGAFMTAPEVLDADRTVAEDADLRLKAAAVTPRATANQFLRIDYRG
NTGLFPMDSVIAEYRWYDSLWVTPTEVTRTGYTFATRWRIQIPLTFSGASTQFD
VTGQSRPQLPEGFTRAYRWAGDGSATEFRGAERVDRAVVRSDTVAPTDQAAAE
KAGAROLLILNDGKPKFDPWADLPEAAPLVAISLGTDDSLARLARFRGATTLRVVS
HPVPRYAYDLVRHHDGAVRPDPVRYAPAGELARVDDTFRDTSGRAVEYRQDLSLQ
LOHPLTVQNRAGSLTSWVTADDVDRVWVSFASRDLQGRGVARSYERSTRTWFAP
IGHPRLLSTGSGFPRAGDNTSVMTAMGSDSGHAGVYVWADGDTSRISLYOGGE
LLGDENVERIVMGGLSPGPKPYRLVLEGSRLNRPDRYSTRTVWDTSATTDPTRL
TLPVLVDVAVAVDLSGRAHRTLTVAHLEGAAGAIAIRATVATVEVSYDDCATWH
RTALRKSADGWTARLDAPGRARYASLRTAKDTGEGVGTQLIRAFGLR"
complement(4374..4613)
/gene="SC8all.04c"
/note="Pfam match to entry PF02225 PA, PA domain, score
19.00, E-value 0.058"
```

Query Match		20.9%; Score 85.2; DB 1; Length 32704;
Best Local Similarity		54.0%; Pred. No. 0.011;
Matches	174; Conservative	0; Mismatches 148; Indels 0; Gaps 0;
QY	38	ttccggtgagctccgatacaggtgcgcggtatcgctatgccatccggtgagcttccacc 97
Db	2896	TTCCGGCGAGTTCGTTACGATCCCGACCCCTCCTCGTCACGATCCGCTTCGCC 2955
QY	98	ttcccgagatgcctctgacctggcggttgcgcgcgagctgctgctggagcggtca 157
Db	2956	OCGAGGGCGCCCCACCGTCACTGGCCATGTGGCGCGTGAACCTGTGACGAGGCCCTGC 3015
QY	158	acagcccgagcgagcgatgtgcacatcgcccgacgagcccgagggcctcgag 217
Db	3016	GCACACGAGCGGCTGGGGACGTCCAGGTGTGGCCGACACCCACCCACCGGGAGA 3075
QY	218	atgtccacatccggtccaggttcgagcgacgtgcgtctgcggcgagggcgac 277
Db	3076	CGCGTGCCTCAGGTCAACGACACGCTGACATAGCCATCTTACGCTGCCCGTCCGG 3135
QY	278	cqctggtggtcttcctcagcgagcagcaagctcgtccgctcgtccgagcagcacgc 337
Db	3136	AACTGGAGGAGTGGAATTACCCACGTAACCTGCACGTCCCGCGGACGGAGTCTGCG 3195
QY	338	tgggtgacttcagcgcaacct 359
Db	3196	GGCTCGGAACGACGCGCTTCCT 3217

RESULT 15  
SC5F2A/c SC5F2A 37245 bp DNA linear BCT 05-MAY-2000  
LOCUS Streptomyces coelicolor cosmid 5F2A.  
DEFINITION AL049587  
ACCESSION AL049587.2 GI:7768282

KEYWORDS ABC transporter; amino acid permease; BCT family; dimethylaminohydrolase; D-amino acid oxidase; ddah, dimethylarginine carboxylaminohydrolase; dehydrogenase; efflux protein; endonuclease; enoyl CoA hydratase; fpp isomerase; metalloproteinase; oxidoreductase; racemase; regulator; thiolase; transferase.

SOURCE Streptomyces coelicolor A3(2).

ORGANISM Streptomyces coelicolor A3(2)

REFERENCE 1 (bases 1 to 37245)

AUTHORS Redenbach,M., Kieser,H.M., Denapaita,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.

TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)

MEDLINE 97000351

REFERENCE 2 (bases 1 to 37245)

AUTHORS Oliver,K. and Harris,D.

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 37245)

AUTHORS Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.

TITLE Direct Submission

JOURNAL Submitted (07-APR-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT On May 10, 2000 this sequence version replaced gi:458464.

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.

URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>

CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid 5F2A.

Location/Qualifiers

1. 37245

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="cosmid 5F2A"

1. .105

/note="Nominal overlap with cosmid 4C6"

complement(83..859)

/gene="SC5F2A.01c"

/note="ddah"

complement(83..859)

/gene="SC5F2A.01c"

FEATURES

source

misc\_feature

gene

CDS



Db	6066	CCACGCCCCGAGGGACCGCGGTGATCCACGTCCGCTCGCGGGCGAGCTGC	CGCGGTTCCT	6007
Qy	294	gcacgggacgacaagctcgtgcgcctcgccagagacacacgctggtgacttcgacg	353	
Db	6006	CCAGCGCCGCGGAGCTGGTCCGGTGGGTCTGAGCACCTCCAGCTGGACCTGACCA	5947	
Qy	354	caactgtgaggacgcactgggcgcgcatc	381	
Db	5946	CGACCTTGCGGAGCTGATCGCGGGGAGC	5919	

Search completed: July 18, 2002, 11:21:43  
Job time: 13713 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:35:47 ; Search time 646.57 Seconds  
(without alignments)  
1080.756 Million cell updates/sec

Title: US-09-749-185-2

Perfect score: 407

Sequence: 1 atagcttctctgtctccga.....gaggagcagaacgcggctg 407

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	407	100.0	407	21	AAZ49728 Streptomyces grise
2	407	100.0	438	21	AAZ49727 Streptomyces grise
3	298.2	73.3	407	21	AAZ49731 Streptomyces netro
4	259.8	63.8	407	21	AAZ49730 Streptomyces golde
5	239	58.7	407	21	AAZ49729 Streptomyces albus
6	75.4	18.5	45624	22	AAF88315 S. spinosa DNA fra
7	75.4	18.5	50000	22	AAF88312 S. spinosa DNA fra
8	54.8	13.5	80161	20	AAZ21501 DNA fragment of Sa
9	54.4	13.4	1549	16	AAQ88151 Xanthomonas L-glut

c	10	53.6	13.2	438	21	AAZ49727 Streptomyces grise
	11	53.4	13.1	1981	20	AAZ49729 Pseudomonas alpha-
	12	53.4	13.1	1981	20	AAZ49729 Pseudomonas alpha-
	13	53.4	13.1	1981	20	AAZ49729 Pseudomonas alpha-
	14	53.4	13.1	1981	20	AAZ49729 Pseudomonas alpha-
	15	53.4	13.1	1981	20	AAZ49729 Pseudomonas alpha-
	16	53.4	13.1	1981	20	AAZ49729 Pseudomonas alpha-
	17	53.4	13.1	1981	20	AAZ49729 Pseudomonas alpha-
	18	51.6	12.7	1313	22	AAF74867 Leishmania major P
	19	51.6	12.7	65140	22	AAZ49727 Streptomyces grise
	20	51.6	12.7	125401	22	AAZ49727 Streptomyces grise
	21	50.4	12.4	42000	21	AAZ49727 Streptomyces grise
	22	50.4	12.4	63164	21	AAZ49727 Streptomyces grise
	23	50.4	12.4	109519	22	AAZ49727 Streptomyces grise
	24	49.6	12.2	2900	13	AAZ49727 Streptomyces grise
	25	49.6	12.2	58857	21	AAZ49727 Streptomyces grise
	26	49.6	12.2	77536	21	AAZ49727 Streptomyces grise
	27	49	12.0	29879	14	AAZ49727 Streptomyces grise
	28	49	12.0	154746	24	AAZ49727 Streptomyces grise
	29	48.8	12.0	16767	22	AAZ49727 Streptomyces grise
	30	48.8	12.0	50000	22	AAZ49727 Streptomyces grise
	31	48.8	12.0	50000	22	AAZ49727 Streptomyces grise
	32	48.8	12.0	109519	22	AAZ49727 Streptomyces grise
	33	48.6	11.9	407	21	AAZ49727 Streptomyces grise
	34	48.6	11.9	3162	23	AAZ49727 Streptomyces grise
	35	48	11.8	1104	22	AAZ49727 Streptomyces grise
	36	48	11.8	2064	14	AAZ49727 Streptomyces grise
	37	47.4	11.6	1161	15	AAZ49727 Streptomyces grise
	38	47.4	11.6	117213	19	AAZ49727 Streptomyces grise
	39	47.2	11.6	2488	22	AAZ49727 Streptomyces grise
	40	47.2	11.6	125401	22	AAZ49727 Streptomyces grise
	41	46.6	11.4	27541	22	AAZ49727 Streptomyces grise
	42	46.4	11.4	1542	21	AAZ49727 Streptomyces grise
	43	46.4	11.4	1542	22	AAZ49727 Streptomyces grise
	44	46.4	11.4	11604	22	AAZ49727 Streptomyces grise
	45	46.4	11.4	15079	16	AAZ49727 Streptomyces grise

ALIGNMENTS

RESULT 1

AAZ49728  
ID AAZ49728 standard; DNA; 407 BP.

XX AAZ49728;

AC AAZ49728;

XX 18-APR-2000 (first entry)

DT Streptomyces griseus ssgA gene-2.

DE ssgA; liquid culture; filamentous bacteria; secondary metabolite;

KW mycelium; antibiotic, antitumor agent; immunosuppressive agent;

KW hypochlosterolemic agent; enzyme inhibitor; antimigraine agent;

KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23345;

KW bioinsecticide; receptor agonist; antagonist; biomass; ds.

OS Streptomyces griseus.

XX Key Location/Qualifiers

FT mat\_peptide 1..405

FT /\*tag= a

FT /label= SsgA\_protein

XX WO200000613-A1.

XX 06-JAN-2000.

XX 25-JUN-1999; 99WO-NL00395.

XX 26-JUN-1998; 98EP-0202148.

XX (UYLE-) RIJKSUNIV LEIDEN.





D	b	121	tgggcgctcgccggagctgctgctgcagcggatcaacccgcccggagcgagcgagcgagc	180
Q	y	181	gtgcacatcgcccgaccgagcccgagggcctcgagatgtccacatccgggtcccaggtc	240
D	b	181	gtccacatcgcccgaccgagcccgagggcctgctcgagcgttccatccgggtcccaggtg	240
Q	y	241	ggcgcgagcgtgcgctgttctccggcggggagcagcgaccgctggtgcttctcgaccgg	300
D	b	241	ggcgcgagcgtgcgctgttctccgctcgagcgcccgccgctgctcctcgaccgc	300
Q	y	301	acggacaagctcgctgcgcctgcgcagagagcacacgctgggtgacttcgacggaacctg	360
D	b	301	acggacaagctcgctgcgcctgcgcagagagcacacgctgggtgacttcgagagacgctg	360
Q	y	361	gagcagcactggccgcatactctccgcgagagagcagaacccggctg	407
D	b	361	gagcgcgctcggaagatcctccgcgagagagcagaacccggctg	407
RESULT 4				
AAZ49730				
ID AAZ49730 standard; DNA; 407 BP.				
X	X			
X	X	AC	AAZ49730;	
X	X	DT		
X	X	18-APR-2000	(first entry)	
X	X	Streptomyces goldeniensis ssGA gene.		
X	X			
K	W	ssGA; liquid culture; filamentous bacteria; secondary metabolite;		
K	W	mycelium; antibiotic, antitumour agent; immunosuppressive agent;		
K	W	hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;		
K	W	herbicide; antiparasitic agent; ruminant growth promoter; ATCC 21386;		
K	W	bioinsecticide; receptor agonist; antagonist; biomass; ds.		
X	X			
O	S	Streptomyces goldeniensis.		
X	X			
F	H	Key	Location/Qualifiers	
F	T	mat_peptide	1..405	
F	T		/*tag= a	
F	T		/label= SsgA_protein	
X	X			
P	N	WO200000613-A1.		
X	X			
P	D	06-JAN-2000.		
X	X			
P	F	25-JUN-1999;	99WO-NL00395.	
X	X			
P	R	26-JUN-1998;	98EP-0202148.	
X	X			
P	A	(UYLE-) RIJKSUNIV LEIDEN.		
P	A	(NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.		
X	X			
P	I	Van Wezel GP, Kraal B, Luiten RGM;		
X	X			
D	R	WPI; 2000-147269/13.		
D	R	P-PSDB; AAY44651.		
X	X			
P	T	Reducing branching and enhancing fragmentation in filamentous		
X	X	microorganisms used to improve their liquid culturing properties		
X	X	- Disclosure; Fig 5; 60pp; English.		
X	X			
C	C	The present sequence is a DNA (ATCC 21386) encoding S. goldeniensis SsgA		
C	C	protein. SsgA reduces branching and fragment septation and enhances		
C	C	fragmentation of mycelium in liquid culture resulting in lower viscosity		
C	C	of culture broths. Filamentous bacteria can be transformed with ssGA		
C	C	gene-containing plasmid to enhance the production of secondary		
C	C	metabolites such as, antibiotics, antitumour agents, immunosuppressive		
C	C	agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine		
C	C	agents, herbicides, antiparasitic agents, ruminant growth promoters,		
C	C	bioinsecticides, receptor agonists and antagonists and biomass.		
X	X			

$\ddot{O}$ 

Query Match	58.7%;	Score 239;	DB 21;	Length 407;
Best Local Similarity	74.2%;	Pred. No. 1.1e-36;		
Matches 302; Conservative	0;	Mismatches 105;	Indels	0;
			Gaps	0;

DT 18-APR-2000 (first entry)

ssgA; liquid culture; filamentous bacteria; secondary metabolite; mycelium; antibiotic, antitumour agent; immunosuppressive agent; hypcholesterolaemic agent; enzyme inhibitor; antimigraine agent; herbicide; antiparasitic agent; ruminant growth promoter; ATCC 30163; blinsecticide; receptor agonist; antagonist; biomass; ds.

WO20000613-A1

XX  
PF  
25-JUN-1999. 99WO-NR.00395

XX  
DB 36-TUN-1000. 00EP-0202140

XX  
PAGE - PLTVCOUNTY EIDEN

PA (NEW-) NEDERLANDSE ORG WETENSCHAPPEL.IJK ONDERZO

PT Van Wezel GP. Kraai B. Luiten BGM.

XX  
NR  
WPT: 2000-147269/13

XX Reducing branching and enhancing fragmentation in filamentous  
PT microorganisms used to improve their liquid culturing properties  
PT  
XX  
PS Disclosure: Fig 5: 60pp: English.

The present sequence is a DNA (ATCC 3004) encoding *S. albus* G SsgA protein. SsgA reduces branching and fragment septation and enhances fragmentation of mycelium in liquid culture resulting in lower viscosity of culture broths. Filamentous bacteria can be transformed with ssgA gene-containing plasmid to enhance the production of secondary metabolites such as, antibiotics, antitumour agents, immunosuppressive agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine agents, herbicides, antiparasitic agents, ruminant growth promoters, bioinsecticides, receptor agonists and antagonists and biomass.

Sequence 407 BP: 57 A; 150 C; 139 G; 61 T; 0 other;

Query Match 58.7%; Score 239; DB 21; Length 407;

Best Local Similarity 74.28; Pred. No. 1.1e-36;  
Matches 302; Conservative 0; Mismatches 105;  
Indels 0; Gaps 0;

Qy 1 atgagcttcctcgctccgagagctctcggtccggtatccggtggagctccgataccgag 60  
|||||

pb 1 atgagcttcctcgctccgagagctccggtccggtatccggtggagctccgataccgag 60  
|||||

Qy 61 gtccgcgatccggtatgccatccggatgacgttccacttcccgccgatgccctgtgacc 12

[illegible][illegible]

**Oy** 241 ggcgggacgctgcgtgttccgggagcggcacgcgtgttggcgttcctcgaaccg 300  
||| ||| | ||||| ||||| ||| ||| ||||| ||| ||||| |||  
**Pb** 241 ggcctccgaacagaggcgtcttccgcgttcggcaagcgcgcgtgctcgccttcctcgaaccg 300

**Qy** 301 acggacaagctcgtgcccgcctcgccaggacacacgcctgggtgaacttcagcggcaacctg 360  
||| ||| | ||| ||| ||| | | ||| ||| ||| |||  
**Dh** 301 acgaaccagaacctatcgctcggcagcagcggcgacacgcccaacttcacagcacctc 360

[illegible]

## RESULT 6

AAF88315/c  
ID AAF88315 standard: DNA: 45624 BP.

AA AF88315:

DT 28-AUG-2007 (first entry)

XX DE S enlrosa DNA fragment SEO TD 4

XX  
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;  
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
KW macrolide; insecticide; do

XX

US  
XX  
saccharopolysp

PN  
XX  
DET995 / 268-A1.

08-MAR-2001.  
PD  
yy



```

Db 24831 gcggactcttgaacacacacacagctggtcgaaacctggtgatgaacaccggtggtg 24890
Qy 343 gacttcgacgcaacctgtaggagcagcactggcgcgatccct 383
Db 24891 aacgtcgacgaggtgctgagccagctgctctcgacacaacct 24931

RESULT 8
ID AAZ21501 standard; DNA; 80161 BP.
XX AC AAZ21501;
XX DT 01-DEC-1999 (first entry)
XX DE DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.
XX KW Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW microtubules; arachnid; nematode; insect; polyketide; polyketide synthase;
KW PKS; extender module; initiator module; acyl transferase domain; AT;
KW acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW insecticide; ss.
XX OS Saccharopolyspora spinosa.
XX FH Location/Qualifiers
FT CDS complement (1135..1971)
FT FT /*tag= a
FT FT /product= ORFL16
FT FT /note= "Protein involved in transcription control"
FT CDS 2024..2791
FT FT /*tag= b
FT FT /product= ORFL15
FT FT /note= "keto acyl reductase"
FT CDS complement (3416..4165)
FT FT /*tag= c
FT FT /product= spns
FT FT /note= "Spinosyn biosynthesis protein S"
FT CDS complement (4168..5325)
FT FT /*tag= d
FT FT /product= spnr
FT FT /note= "Spinosyn biosynthesis protein R"
FT CDS complement (5363..6751)
FT FT /*tag= e
FT FT /product= spnQ
FT FT /note= "Spinosyn biosynthesis protein Q"
FT CDS 7083..8450
FT FT /*tag= f
FT FT /product= spnp
FT FT /note= "Spinosyn biosynthesis protein P"
FT CDS 8967..10427
FT FT /*tag= g
FT FT /product= SpnO
FT FT /note= "Spinosyn biosynthesis protein O"
FT CDS 10436..11434
FT FT /*tag= h
FT FT /product= spnN
FT FT /note= "Spinosyn biosynthesis protein N"
FT CDS complement (11530..12492)
FT FT /*tag= i
FT FT /product= spnM
FT FT /note= "Spinosyn biosynthesis protein M"
FT CDS complement (12696..13547)
FT FT /*tag= j
FT FT /product= spnL
FT FT /note= "Spinosyn biosynthesis protein L"
FT CDS complement (13592..14785)
FT FT /*tag= k
FT FT /product= spnK
FT FT /note= "Spinosyn biosynthesis protein K"
FT CDS complement (14799..16418)

```

```

FT FT /*tag= l
FT FT /product= spnJ
FT FT /note= "Spinosyn biosynthesis protein J"
FT CDS 16556..17743
FT FT /*tag= m
FT FT /product= spnI
FT FT /note= "Spinosyn biosynthesis protein I"
FT CDS complement (17749..18501)
FT FT /*tag= n
FT FT /product= spnH
FT FT /note= "Spinosyn biosynthesis protein H"
FT CDS complement (18541..19713)
FT FT /*tag= o
FT FT /product= spnG
FT FT /note= "Spinosyn biosynthesis protein G"
FT CDS 20168..20995
FT FT /*tag= p
FT FT /product= spnF
FT FT /note= "Spinosyn biosynthesis protein F"
FT CDS 21111..28898
FT FT /*tag= q
FT FT /product= spnA
FT FT /note= "Spinosyn biosynthesis protein A"
FT FT /transl_except= (Pos:26940..26942, aa:Pro)
FT misc_feature 21126..22379
FT FT /*tag= r
FT FT /note= "Beta-ketosynthase domain (KSI): part of the
FT FT initiator module"
FT misc_feature 22692..23669
FT FT /*tag= s
FT FT /note= "Acyl transferase domain (AT1): part of the
FT FT initiator module"
FT misc_feature 23793..24041
FT FT /*tag= t
FT FT /note= "Acyl carrier protein domain (ACPI): part of the
FT FT initiator module"
FT misc_feature 24102..25349
FT FT /*tag= u
FT FT /note= "Beta-ketosynthase domain (KSI): part of
FT FT extender module 1"
FT misc_feature 25683..26684
FT FT /*tag= v
FT FT /note= "Acyl transferase domain (AT1): part of
FT FT extender module 1"
FT misc_feature 27582..28121
FT FT /*tag= w
FT FT /note= "Beta-ketoreductase domain (KR1): part of
FT FT extender module 1"
FT misc_feature 28404..28649
FT FT /*tag= x
FT FT /note= "Acyl carrier protein domain (ACPI): part of
FT FT extender module 1"
FT CDS 38916..35374
FT FT /*tag= y
FT FT /product= spnC
FT FT /note= "Spinosyn biosynthesis protein C"
FT misc_feature 29024..30295
FT FT /*tag= z
FT FT /note= "Beta-ketosynthase domain (KS2): part of
FT FT extender module 2"
FT misc_feature 30629..31621
FT FT /*tag= aa
FT FT /note= "Acyl transferase domain (AT2): part of
FT FT extender module 2"
FT misc_feature 31697..32254
FT FT /*tag= ab
FT FT /note= "Dehydratase domain (DH2): part of extender
FT FT module 2"
FT misc_feature 33035..34072
FT FT /*tag= ac
FT FT /note= "Enoyl reductase domain (ER2): part of
FT FT extender module 2"
FT misc_feature 34082..34621

```

Qy	139	ctcgtcgtgacgggctcaacagcccgagcgcgcgcgcgatgtgcacatcgcgccgacc	198
Db	62	ctcgtggccgacgggctgctgcgcgaggcgcgaagcgatgcggatcgcgcc---ct	118
Qy	199	gagccgagggcctcggagatgtccacatccggctccaggtcggcgcgcgcgcgcgtg	258
Db	119	cgcaggggttttcgggggttgctcgatcgatgagctgcgcgtcggggcaggcctcc	178
Qy	259	tcccggggcgggagcgcacccgtggctggcgttctctcgcgcgcgcgcgcgcgcgc	318
Db	179	tccgaggtgaactgctgaccagcttgcggacttcttgacgcacacctacgcacgtggtcgaa	238
Qy	319	ctcggcccgagcgcacgcgtgggtgacttcgcgcgcgcgcgcgcgcgcgcgcgcgc	360
Db	239	cctggtgatgaacacgggtggatgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	280
 RESULT 9 AAQ88151 ID AAQ88151 standard; DNA; 1549 BP. XX AC AAQ88151; XX DT 02-NOV-1995 (first entry) XX Xanthomonas L-glutamyl tRNA reductase. XX DE Xanthomonas campestris; L-glutamyl tRNA reductase; vitamin B12; KW glutamic acid-1-semialdehyde; 5-aminolevulinic acid (ALA); haem; KW chlorophyll; phycobillin; recombinant production; ds. XX OS Xanthomonas campestris. XX FH Key Location/Qualifiers FT CDS 212..1492 FT /*tag= a XX PN JF07031480-A. XX PD 03-FEB-1995. XX PF 27-JUL-1993; 93JP-0184709. XX PR 27-JUL-1993; 93JP-0184709. XX PA (COSM-) COSMO SOGO KENKYUSHO KK. XX WPI; 1995-109535/15. XX P-PSDB; AAR74619. XX PT DNA fragment encoding L-glutamyl tRNA reductase - derived from PT Xanthomonas sp.; useful for production of 5-amino:levulinic, PT vitamin-B12, chlorophyll, etc. XX PS Claim 3; Pages 7-8; 9pp; Japanese. XX CC AAQ88151 encodes AAR74619 Xanthomonas campestris L-glutamyl tRNA CC reductase. The DNA as part of an expression vector can be used to CC transform a host cell, for the recombinant prodn. of the protein. CC The protein can be used in the prodn. of vitamin B12, glutamic CC acid-1-semialdehyde, 5-aminolevulinic, haem, chlorophyll and XX phycobillin. SQ Sequence 1549 BP; 291 A; 516 C; 514 G; 228 T; 0 other;			

[illegible]



```
Query Match 13.1%; Score 53.4; DB 20; Length 1981;
Best Local Similarity 51.0%; Pred. No. 0.087;
Matches 126; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 117 gacttggcgttcgcccgcgagctgctgctgacgaggtctcaacagccgagcgagcg 176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 582 gatctccgactacctgctcgacactggtccagatggcggtcaagggttgcgctgatgc 641
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 177 cgatgtcacatcgcccgacgagcccgagggcctcgagatgtccacatccgctcca 236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 642 gccaagcacatcagcccgacgagacctgggtgccatcatcgacagcgcaacg 701
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 237 ggtcggcgagaccgtgcgtgttccggcgagcgacgacctggtggcgttctctga 296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 702 cggtcgcgacgacctggttctgaggtgatcgcgccggcgagcggtgca 761
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 297 ccgagcgacaagctgcgcgtcgccgagcagcgacacacgctgggtgacttcgagc 356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 762 gccagccagctacttcgggtcgccggcgaggtcaggtgacgaggttcgctacgg 821
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 357 cctggag 363
    |||||
Db 822 caaggag 828
    |||||

RESULT 12
AAX37294
ID AAX37294 standard; DNA; 1981 BP.
XX
AC AAX37294;
XX
DT 27-JUL-1999 (first entry)
XX
DE Pseudomonas alpha-amylase variant encoding DNA (FERM BP-6119).
XX
KW Pseudomonas; maltopentaose; alpha-amylase; enzyme; variant; ss.
XX
OS Pseudomonas sp.
XX
OS Synthetic.
XX
PN JP11123081-A.
XX
PD 11-MAY-1999.
XX
PF 21-OCT-1997; 97JP-0305071.
XX
PR 21-OCT-1997; 97JP-0305071.
XX
PA (NORQ ) NORINSUISANSHO SHOKUHN SOGO.
XX
DR WPI; 1999-340517/29.
XX
DR P-PSDB; AAY17905.
XX
PT New alpha-amylase gene - useful for production of malto-pentaose
XX
PS Claim 5; Page 9-11; 23pp; Japanese.
XX
CC Sequences AAX37293-298 represent nucleic acids encoding alpha-amylase
    variants (AAY17904-17909). The variants are derived from the
    CC Pseudomonas sp. KO-8940-derived maltopentaose-forming amylase (Shida, O.
    CC et al, Biosci. Biotech. Biochem. Vol. 56, 76-80, 1992 ). The variants
    CC have a replacement of the 57th or the 130th amino acid residue in the
    CC sequence as compared to the wild-type. The sequences are deposited under
    CC the accession numbers FERM BP-6116, 6119, 6115, 6117, 6118 or 6114. The
    CC variants can be used to transform plants to produce maltopentaose. The
    CC enzymes prepared from the alpha-amylase variants produce maltopentaose
    CC in a high yield.
XX
SQ Sequence 1981 BP; 343 A; 679 C; 678 G; 281 T; 0 other;

Query Match 13.1%; Score 53.4; DB 20; Length 1981;
Best Local Similarity 51.0%; Pred. No. 0.087;
Matches 126; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 117 gacttggcgttcgcccgcgagctgctgctgacgaggtctcaacagccgagcgagcg 176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 582 gatctccgactacctgctcgacactggtccagatggcggtcaagggttgcgctgatgc 641
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 177 cgatgtcacatcgcccgacgagcccgagggcctcgagatgtccacatccgctcca 236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 642 gccaagcacatcagcccgacgagacctgggtgccatcatcgacagcgcaacg 701
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 237 ggtcggcgagaccgtgcgtgttccggcgagcgacgacctggtggcgttctctga 296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 702 cggtcgcgacgacctggttctgaggtgatcgcgccggcgagcggtgca 761
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 297 ccgagcgacaagctgcgcgtcgccgagcagcgacacacgctgggtgacttcgagc 356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 762 gccagccagctacttcgggtcgccggcgaggtcaggtgacgaggttcgctacgg 821
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 357 cctggag 363
    |||||
Db 822 caaggag 828
    |||||

RESULT 13
AAX37295
ID AAX37295 standard; DNA; 1981 BP.
XX
AC AAX37295;
XX
DT 27-JUL-1999 (first entry)
XX
DE Pseudomonas alpha-amylase variant encoding DNA (FERM BP-6115).
XX
KW Pseudomonas; maltopentaose; alpha-amylase; enzyme; variant; ss.
XX
OS Pseudomonas sp.
XX
OS Synthetic.
XX
PN JP11123081-A.
XX
PD 11-MAY-1999.
XX
PF 21-OCT-1997; 97JP-0305071.
XX
PR 21-OCT-1997; 97JP-0305071.
XX
PA (NORQ ) NORINSUISANSHO SHOKUHN SOGO.
XX
DR WPI; 1999-340517/29.
XX
DR P-PSDB; AAY17906.
XX
PT New alpha-amylase gene - useful for production of malto-pentaose
XX
PS Claim 6; Page 11-14; 23pp; Japanese.
XX
CC Sequences AAX37293-298 represent nucleic acids encoding alpha-amylase
    variants (AAY17904-17909). The variants are derived from the
    CC Pseudomonas sp. KO-8940-derived maltopentaose-forming amylase (Shida, O.
    CC et al, Biosci. Biotech. Biochem. Vol. 56, 76-80, 1992 ). The variants
    CC have a replacement of the 57th or the 130th amino acid residue in the
    CC sequence as compared to the wild-type. The sequences are deposited under
    CC the accession numbers FERM BP-6116, 6119, 6115, 6117, 6118 or 6114. The
    CC variants can be used to transform plants to produce maltopentaose. The
    CC enzymes prepared from the alpha-amylase variants produce maltopentaose
    CC in a high yield.
XX
SQ Sequence 1981 BP; 342 A; 679 C; 678 G; 282 T; 0 other;
```







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:24:25 ; Search time 165.56 Seconds  
(without alignments)  
603.847 Million cell updates/sec

Title: US-09-749-185-2  
Perfect score: 407  
Sequence: 1 atgagcttctctgtctccga.....gaggagcagaacgcggctg 407

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.8	13.5	80161	3	US-09-036-987A-1
2	54.8	13.5	80161	4	US-09-370-700-1
3	53.4	13.1	1981	3	US-09-017-706-3
4	53.4	13.1	1981	3	US-09-017-706-4
5	53.4	13.1	1981	3	US-09-017-706-5
6	53.4	13.1	1981	3	US-09-017-706-6
7	53.4	13.1	1981	3	US-09-017-706-7
8	53.4	13.1	1981	3	US-09-017-706-8
9	49.6	12.2	2900	1	US-08-034-650-9
10	49.6	12.2	2900	1	US-08-449-015-9
11	49	12.0	1155	2	US-08-387-942C-7
12	49	12.0	11219	1	US-07-642-734C-1
13	49	12.0	11219	3	US-08-439-009A-1
14	48	11.8	2064	1	US-08-343-428-1
15	46.4	11.4	1542	4	US-09-385-028-16
16	46.4	11.4	11604	4	US-09-385-028-13
17	46.4	11.4	15079	4	US-09-385-028-1
18	46	11.3	2745	1	US-08-363-255-1
19	46	11.3	2745	1	US-08-363-255-13
20	45.8	11.3	1155	2	US-08-387-942C-18
21	45.8	11.3	12588	4	US-08-387-942C-1
22	45.4	11.2	2439	4	US-09-624-693A-18
23	44.8	11.0	1212	6	5212296-17
24	44.8	11.0	1998	6	5212296-8
25	44	10.8	8931	3	US-09-051-019-1
26	44	10.8	20235	1	US-07-642-734C-3
27	44	10.8	20235	3	US-08-439-009A-3

28	43.8	10.8	44377	2	US-08-804-227C-7
29	43.8	10.8	44377	2	US-08-804-198-1
30	43.6	10.7	1683	1	US-07-945-283-3
31	43.6	10.7	1831	6	5215881-1
32	43.6	10.7	1831	6	5215881-3
33	43.6	10.7	8438	1	US-07-945-283-1
34	43.4	10.7	9515	1	US-08-920-812-13
35	43.4	10.7	9515	1	US-08-920-827-13
36	43.4	10.7	9515	1	US-08-921-177-13
37	43.4	10.7	9515	1	US-08-362-577C-13
38	43.4	10.7	9515	2	US-08-920-828-13
39	43	10.6	20235	1	US-07-642-734C-3
40	43	10.6	20235	3	US-08-439-009A-3
41	43	10.6	71989	4	US-09-443-501A-2
42	42.4	10.4	1294	3	US-09-025-691-2
43	42.4	10.4	44377	2	US-08-804-227C-7
44	42.4	10.4	44377	2	US-08-804-198-1
45	42.2	10.4	999	2	US-08-761-277A-46

ALIGNMENTS

RESULT 1  
US-09-036-987A-1  
; Sequence 1, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Treadway, Patti J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; TITLE OF SEQUENCES: 39  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow Agrosciences LLC Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036,987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R  
; REGISTRATION NUMBER: 28,479  
; REFERENCE/DOCKET NUMBER: 50,608  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317)337-4816  
; TELEFAX: (317)337-4847  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 80161 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-09-036-987A-1

Query Match 13.5%; Score 54.8; DB 3; Length 80161;



Qy 117: gacctgggcgttcggccgcgcgaactgctgctgacggctcaacagcccgagcggcgacgg 176



APPLICANT: VISSEER, Christiaan  
TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE  
TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1615 L. Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/034.650  
FILING DATE: 03-JUL-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/727,235  
FILING DATE: 03-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kokulis, Paul N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: PNK/5970/91731  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2900 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 483..1556  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1559..2617  
US-08-034-650-9

APPLICANT: VISSEER, Christiaan  
TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE  
TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1615 L. Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449.015  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/727,235  
FILING DATE: 03-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kokulis, Paul N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: PNK/5970/91731  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2900 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 483..1556  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1559..2617  
US-08-449-015-9

Query Match 12.2%; Score 49.6; DB 1; Length 2900;  
Best Local Similarity 50.9%; Pred. No. 0.071;  
Matches 118; Conservative 0; Mismatches 114; Indels 0; Gaps 0;  
Qy 112 cctgtgacctggcggttcggcgagctgtgtgagcggtctcaacagcccgagcggc 171  
Db 2081 CCGCGCGCATGACGTCGCGTCGATCAGCGCGCGGCGTGGCGGACCGACGCTCGGC 2140  
Qy 172 gacggcgatgtgcacatcgcccgaccgagcccgagggcctcggagatgtccacatccgg 231  
Db 2141 GAGTGGCGGACGCGTTCGCGGACGACGCGCGCGGCGCATGACGCTCGAACGG 2200  
Qy 232 ctccaggtcgcgcgagccgtgctgttccggggcgagcgccacccgctgtgtggttc 291  
Db 2201 ATCCGGATCGCCACGACACACCGCTGAGCCCTGAGCAGAGGCGCGGCTTCCGCG 2260  
Qy 292 ctccagcgagcggaacgctgctgcgtcgccagggagcagcagctgggtg 343  
Db 2261 CTCGACGCGCAGCTGACGCGCGGACGAGCGCGCGGCGGCGCTGCATG 2312

RESULT 10  
US-08-449-015-9  
; Sequence 9, Application US/08449015  
; Patent No. 5804409  
; GENERAL INFORMATION:  
; APPLICANT: BOS, Jannetje W.  
; APPLICANT: FRENKEN, Leon G.  
; APPLICANT: VERRIPS, Cornelis T.

Query Match 12.2%; Score 49.6; DB 1; Length 2900;  
Best Local Similarity 50.9%; Pred. No. 0.071;  
Matches 118; Conservative 0; Mismatches 114; Indels 0; Gaps 0;  
Qy 112 cctgtgacctggcggttcggcgagctgtgtgagcggtctcaacagcccgagcggc 171  
Db 2081 CCGCGCGCATGACGTCGCGTCGATCAGCGCGGCGGCGTGGCGGACCGACGCTCGGC 2140  
Qy 172 gacggcgatgtgcacatcgcccgaccgagcccgagggcctcggagatgtccacatccgg 231  
Db 2141 GAGTGGCGGACGCGTTCGCGGACGACGCGCGCGGCGCATGACGCTCGAACGG 2200  
Qy 232 ctccaggtcgcgcgagccgtgctgttccggggcgagcgccacccgctgtgtggttc 291  
Db 2201 ATCCGGATCGCCACGACACACCGCTGAGCCCTGAGCAGAGGCGCGGCTTCCGCG 2260  
Qy 292 ctccagcgagcggaacgctgctgcgtcgccagggagcagcagctgggtg 343  
Db 2261 CTCGACGCGCAGCTGACGCGCGGACGAGCGCGCGGCGGCGCTGCATG 2312

RESULT 11  
US-08-387-942C-7  
; Sequence 7, Application US/08387942C  
; Patent No. 5939289  
; GENERAL INFORMATION:  
; APPLICANT: ERTESVAG, HELGA





```
; OTHER INFORMATION: acyltransferase 2 domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5574..6125
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6369..6626
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain 2 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6678..11219
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6678..8066
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8262..9305
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9906..10454
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10707..10964
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 2"
US-07-642-734C-1
```

```
Query Match 12.08; Score 49; DB 1; Length 11219;
Best Local Similarity 50.28; Pred. No. 0.088;
Matches 147; Conservative 0; Mismatches 145; Indels 1; Gaps 1;

Qy 60 ggtcgccgcatcgatgccatcccgatgacgttccacgttccggtgagccctgtgac 119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6143 GCCACCGCGCTTCGGCCATCCGCTCCCGCCACGCTCCCGCCACGCGGCT-CG 6085

Qy 120 ctggcggttcggccgagctgctgtgagcggtgctcaacagcccgagcgcgagcgga 179
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6084 CGGGAGTCCGTCCGCTCCGCGCTGCTGGCGGAGGCGGTGAGGTTGCCGGCG 6025

Qy 180 tgtcacatcgccgacccgagccgagccgagccgagccgagccgagccgagccgagcc 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6024 CGTAGCCCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5965

Qy 240 cggcgagccgctgctgttcggttcggtgagccgagccgagccgagccgagccgagcc 299
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5964 CGGTAGTTCAGTCGCGCGCTGAGTTCGTCAGGTTCCGCGCGCGCGCGCGCGCG 5905

Qy 300 gacggacagctcgttcgctcgttcggtgagccagcagcagcagcagcagcagcagc 352
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5904 GACTTCCCGCTCGATGCTCGCTCGCGGTGAGGTTCCACGCGTCCGCTCGT 5852
```

## RESULT 13

```
US-08-439-009A-1/c
; Sequence 1, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
```

```
; TITLE OF INVENTION: Specific Polyketides
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven F. Weinstock
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 2338
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..6659
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
OTHER INFORMATION: MODULE 1"
OTHER INFORMATION: /label= FUNCTION
FEATURE:
NAME/KEY: CDS
LOCATION: 744..11219
OTHER INFORMATION: /function= "gene= "eryA""
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
OTHER INFORMATION: 6-deoxyerythronolide B""
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..1868
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1998..2198
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2250..3626
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3831..4811
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase 2 domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5574..6125
OTHER INFORMATION: /function= "approximate span of
```



Search completed: July 18, 2002, 11:24:51  
Job time: 10881 sec

US-09-385-028-16  
; Sequence 16, Application US/09385028  
; Patent No. 6232106  
; GENERAL INFORMATION:  
; APPLICANT: Susan E. Jensen  
; APPLICANT: Kwamena A. Aidoo  
; APPLICANT: Ashish S. Paradkar  
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Ciavulanic  
; Patent No. 6232106  
; TITLE OF INVENTION: Acid Biosynthesis  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC  
; STREET: The Jenifer Building, 400 Seventh Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/385,028  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/790,462  
; FILING DATE: 29-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: D. Douglas Price  
; REGISTRATION NUMBER: 24,514  
; REFERENCE/DOCKET NUMBER: 1418/P57452US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 638-6666  
; TELEFAX: (202) 39305350  
; TELEX: RCA 248593 IDEA UR  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1542 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-385-028-16

Query Match 11.4%; Score 46.4; DB 4; Length 1542;  
Best Local Similarity 49.7%; Pred. No. 0.29;  
Matches 147; Conservative 0; Mismatches 146; Indels 3; Gaps 1;  
Qy 40 ccggtgagtcgcatacagaggtcgccgacgtatgcctccatccggtacgttccacctt 99  
Db 121 CAGGGGAGCGCTCGCTCGCGGACCGCCCTGTCACGCCCGCTCGGTCCGCCGCGCGC 180  
Qy 100 cccggcagtcgctgacgtggtggtggtggtggtggtggtggtggtggtggtggt 159  
Db 181 GCGGTGGCGCGCTCCCTCACCAGCGCGCGCCACACCGCGGTGCTCGCGGTGAGATCTAC 240  
Qy 160 agcccgagcgcgagcgagtgatgcacatcgccgacgagcgagcgagcgagcgagcgag 219  
Db 241 AACCGGAGAGACTCTCTCGGTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 300  
Qy 220 gtccacatccggctccaggtcgccgagcggtcgctgttccgggagcgagcgagcgag 279  
Db 301 CTGCTCTCGCGCTCGTGAACGCTATGACCTGATGCTTCGGCTGCTGAACGG---G 357  
Qy 280 ctggtgggttcctcgacgagcgagcaagctcgtgcgctcgccgagcgagcgagcgag 335  
Db 358 CGCTTCGCGACCGTGGTGGCGAGCGGGGACCGGGTCTCTGCTCGCGACCGACCGC 413

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 10:12:44 ; Search time 4821.96 Seconds  
(without alignments)  
1139.217 Million cell updates/sec

Title: US-09-749-185-2  
Perfect score: 407  
Sequence: 1 atgagcttcctcgtctccga.....gaggagcagaacgcggctg 407

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_estl:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_in:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57.8	14.2	925	12	CNS0091P
2	57	14.0	925	12	AL053013 Drosophila
3	53	13.0	928	12	AQ856496 nbe0003C
4	52.6	12.9	935	12	CNS006XX
5	52.2	12.8	556	12	AQ851215 LMAJFV1_1
6	52.2	12.8	414	12	AQ847989 LMAJFV1_1
7	50.6	12.4	474	12	AQ851336 LMAJFV1_1
8	50	12.3	755	12	AL192712 Tetraodon
9	49.8	12.2	558	10	C97336 C97336 Rice
10	49.6	12.2	442	12	AQ849196 LMAJFV1_1
11	49.6	12.2	1009	12	CNS010EW
12	49.6	12.2	1129	12	CNS02AGC
13	49.4	12.1	1313	12	AG032885
14	48.8	12.0	437	10	BE400231
15	48.2	11.8	1101	12	CNS017SY
16	48	11.8	390	12	AQ846145 LMAJFV1_1
17	48	11.8	732	12	BH018404 L14841.d

18	47.8	11.7	380	12	AQ846152 LMAJFV1_1
19	47.8	11.7	555	12	AQ851019
20	47.8	11.7	578	12	BH020269
21	47.8	11.7	677	12	BH019193
22	47.8	11.7	705	12	BH019836
23	47.6	11.7	395	12	AQ852043 LMAJFV1_1
24	47.6	11.7	631	9	BB635178
25	47.6	11.7	723	10	B1730817
26	47.4	11.6	609	12	BH019916
27	47.4	11.6	1339	10	BM458211
28	47.2	11.6	615	10	B1529148
29	47.2	11.6	680	10	BG859785
30	47	11.5	912	10	BE418762
31	47	11.5	992	12	AG057280
32	46.8	11.5	452	9	AL588142
33	46.8	11.5	932	12	CNS00720
34	46.8	11.5	1625	12	AG043477
35	46.6	11.4	631	9	AW911065
36	46.6	11.4	828	9	BE040853
37	46.6	11.4	925	12	CNS04EIN
38	46.6	11.4	1085	12	AG152796
39	46.6	11.4	1296	12	AG032939
40	46.4	11.4	332	10	BE597437
41	46.4	11.4	406	12	BH018881
42	46.4	11.4	445	12	AQ849292
43	46.4	11.4	485	12	BH019456
44	46.4	11.4	657	12	AZ131579
45	46.4	11.4	811	10	B1956687

ALIGNMENTS

RESULT 1

CNS0091P

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL053013 GI:4934461  
AL053013.1  
GSS.  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr)  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aron Mammos in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
PI and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
Location/Qualifiers  
1. .925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"









[illegible][illegible]

Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.  
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 755)  
Roest-Crolius, H., Jallou, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fzames, C., Wncker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.  
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 755)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
Location/Qualifiers  
1..755  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG260DB03LP1-end : T7"

BASE COUNT 137 a 238 c 266 g 111 t 3 others  
ORIGIN

Query Match 12.3%; Score 50; DB 12; Length 755;  
Best Local Similarity 49.2%; Pred. No. 18;  
Matches 131; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 91 ttccacattccggcgatgcccctgttgacctgttgccgttcggcgccgagctgctgtgac 150  
|||||  
Db 277 TTCTCGCGCTGCCGGGACAAAGTCCAGGAGCTGATCTCAGCGAGGAGCTGGAGTG 336

Qy 151 gggctaacacggcagcgagcgatgtgcacatggccgacgagcccgagggc 210  
|||||  
Db 337 GAGGACGAGACCTGGTGTAGAGGCGCTGATCAGCTGGGTCAAGGCGCACGAGGCGC 396

Qy 211 ctggagatgtccacatccggctccaggtcgccgagcgacgtgctgttcggcgggg 270  
Db 397 AGGCGGCGAGCTGCCCGCTCTCGCTGCTGCGCTGGCGCTGCTCCCGAGGCC 456

Qy 271 acggcaccgtgtgtggttctctcgacggcagcgacgaagctcgtgcccgtcgccagg 330  
Db 457 TACCTGTGAAGACGGTGGCTCCGAGGAGCTGTGTATGTGCCACAAAGCTGGGCGGGAG 516

Qy 331 cacacgtggtgacttcgacggcaa 356  
Db 517 ATCTGGAGGACGCGCTGCCGTGCAA 542

RESULT 9  
C97336  
LOCUS C97336 Rice callus Oryza sativa mRNA EST 19-OCT-1998  
DEFINITION C97336 Rice callus Oryza sativa cdna clone C60023\_11A, mRNA sequence.  
ACCESSION C97336  
VERSION C97336  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.  
REFERENCE 1 (bases 1 to 558)  
AUTHORS Sasaki, T. and Minobe, Y.  
TITLE Rice cDNA from callus  
JOURNAL Unpublished (1994)

Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.  
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 755)  
Roest-Crolius, H., Jallou, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fzames, C., Wncker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.  
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 755)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
Location/Qualifiers  
1..755  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG260DB03LP1-end : T7"

BASE COUNT 137 a 238 c 266 g 111 t 3 others  
ORIGIN

Query Match 12.3%; Score 50; DB 12; Length 755;  
Best Local Similarity 49.2%; Pred. No. 18;  
Matches 131; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 91 ttccacattccggcgatgcccctgttgacctgttgccgttcggcgccgagctgctgtgac 150  
|||||  
Db 277 TTCTCGCGCTGCCGGGACAAAGTCCAGGAGCTGATCTCAGCGAGGAGCTGGAGTG 336

Qy 151 gggctaacacggcagcgagcgatgtgcacatggccgacgagcccgagggc 210  
|||||  
Db 337 GAGGACGAGACCTGGTGTAGAGGCGCTGATCAGCTGGGTCAAGGCGCACGAGGCGC 396

Qy 211 ctggagatgtccacatccggctccaggtcgccgagcgacgtgctgttcggcgggg 270  
Db 397 AGGCGGCGAGCTGCCCGCTCTCGCTGCTGCGCTGGCGCTGCTCCCGAGGCC 456

Qy 271 acggcaccgtgtgtggttctctcgacggcagcgacgaagctcgtgcccgtcgccagg 330  
Db 457 TACCTGTGAAGACGGTGGCTCCGAGGAGCTGTGTATGTGCCACAAAGCTGGGCGGGAG 516

Qy 331 cacacgtggtgacttcgacggcaa 356  
Db 517 ATCTGGAGGACGCGCTGCCGTGCAA 542

RESULT 9  
C97336  
LOCUS C97336 Rice callus Oryza sativa mRNA EST 19-OCT-1998  
DEFINITION C97336 Rice callus Oryza sativa cdna clone C60023\_11A, mRNA sequence.  
ACCESSION C97336  
VERSION C97336  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.  
REFERENCE 1 (bases 1 to 558)  
AUTHORS Sasaki, T. and Minobe, Y.  
TITLE Rice cDNA from callus  
JOURNAL Unpublished (1994)

Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
305-8602, Japan  
Tel: 81-298-38-7441  
Fax: 81-298-38-7468  
Email: tsasaki@agr.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>  
PROJECT = 'RGP'

FEATURES  
Location/Qualifiers  
1..558  
/organism="Oryza sativa"  
/strain="cultivar Nipponbare, sub\_species Japonica"  
/db\_xref="taxon:4530"  
/clone="C60023\_11A"  
/clone\_lib="Rice callus"  
/note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid."  
71 a 188 c 209 g 86 t 4 others  
BASE COUNT  
ORIGIN

Query Match 12.2%; Score 49.8; DB 10; Length 558;  
Best Local Similarity 51.2%; Pred. No. 19;  
Matches 145; Conservative 0; Mismatches 132; Indels 6; Gaps 1;

Qy 57 cgagtgccgagccgtatgccatccgagtgacgttccacattcccgcgatgccctgt 116  
|||||  
Db 112 CGACCTCGCGCGCGCGCTGACACGCGGCTGTCTTACCGCCCAACAGCGGA 171

Qy 117 gacctggcggttcggcgagctgtgtgacggcggtccacagccgagcgagcgacgg 176  
|||||  
Db 172 GCCGGGTTCTTCGCGCGCGCGCATGAGGACAGCTCGCCAGGGCGCTCGTGGCGTT 231

Qy 177 cgatgtgcacatcgccgagccgagccgagggcctcgagatgtccacatcccggtcca 236  
Db 232 CTACCCGCTCGCGCGCGCGCTCGGCGCTCGA-----CGGCGACGCGCGCTCCAGTCCA 285

Qy 237 ggtcgcgcgagccgtgcgtgttccggggcgagcgaccgctggtggcgttccctga 296  
Db 286 CTGACCCGCGGAGGCGCGTGGTGTTCGCCACGCGGAGGTCGGGCGACTACGCGCTCGACGA 345

Qy 297 ccgagcgacgaagctgcgctcgccgctcgccagagacacacgctg 339  
Db 346 CCTGATGCGCGAGTTCGTGCGCTCGAGAGATCGCGGACCTG 388

RESULT 10  
AQ849196  
LOCUS AQ849196 442 bp DNA linear GSS 25-MAY-2001  
DEFINITION LMAJFV1\_lm43e09.y1 Leishmania major FV1 random genomic library  
Leishmania major genomic clone LMAJFV1\_lm43e09.5, similar to  
contains element VI-chl\_type\_II.2 leishmania repetitive element ;,  
DNA sequence.  
ACCESSION AQ849196  
VERSION AQ849196.1 GI:6053844  
KEYWORDS GSS.  
SOURCE Leishmania major.  
ORGANISM Leishmania major.  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Leishmania.  
REFERENCE 1 (bases 1 to 442)  
AUTHORS Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L., Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A., Ronko, Blisstein, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko, J., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M., Harvey, N., McCann, R., Tsagaris, V., Williams, T., Jackson, Y., Bowers, Y., Swaller, T., Waterston, R., Wilson, R., and Beverley, S.M.  
TITLE A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling  
JOURNAL Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)





Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.

1 (bases 1 to 437)

Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier, S., Dubcovsky, J., Feuillet, C., Gate, M., Graner, A., Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M., Sorrells, M., Warburton, M., and Wenzel, G.

International Triticeae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae

Unpublished (2000)

Contact: Langridge P

Special Research Centre, Basic and Applied Plant Molecular Biology

Waite Campus, University of Adelaide, Glen Osmond SA 5064 AUSTRALIA

Tel: 61 8 8303 7368

Fax: 61 8 8303 7102

Email: plangridge@waite.adelaide.edu.au

International Triticeae EST Cooperative (ITEC)

http://wheat.pw.usda.gov/genome.

## FEATURES

Location/Qualifiers

source

1..437

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db\_xref="taxon:4565"

/clone="AWB001.G03"

/clone\_lib="ITEC AWB Wheat Melotic Stage Library"

/tissue\_type="excised florets"

/dev\_stage="melotic stage no later than metaphase I"

/note="vector: pSport 1 (Life Technologies cat. no. 18248-013); Site 1: SalI; Library constructed in pSport 1.

Directionally cloned using the Superscript Plasmid System

for cDNA synthesis and plasmid cloning. M13 Reverse

sequencing primer used to obtain 5' sequence data. 1.4

Kbp average insert size."

BASE COUNT 56 a 188 c 143 g 50 t

## ORIGIN

Query Match 12.08; Score 48.8; DB 10; Length 437;  
Best Local Similarity 49.2; Pred. No. 27;  
Matches 128; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 132 ccgcagctgctgtagcgggtcaacagccgagcgagcgagcgatgtgcacatcg 191

Db 105 CCGCGCTGACATGATGGAGTCCGGCATGACGCTGGGGCTCGCACCGGCTCCACGGC 164

Qy 192 cccgacgagccgagggcctcgagatgtccacatccggtccaggtcgcgcgagccg 251

Db 165 CGCGACGCGCTCGACCGCTCGCGACCTCTCCGACCGCGCGCTCGCGCGGTGC 224

Qy 252 tgcgctgtccggcggggacgacgctggtggttcttcgacgagcgagcaagct 311

Db 225 CGGGGTGCCACCTCCCTCAAGAGGAGGCGACGCGCGCGCTCGGGATCCCGATGCT 284

Qy 312 cgtccgctcggcagagacacgctgggtgacttcgacgcaacctggagagcact 371

Db 285 CGCGTCCCGACGCGCGCGAGATCCACCTCTCATACGCGCGCGAGGTCGACCC 344

Qy 372 gggcgcatcctcgccgagg 391

Db 345 GGACCTCAACCTCGTCAAGG 364

## RESULT 15

CNS017SY/c

LOCUS CNS017SY 1101 bp DNA linear GSS 26-JUL-1999

DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC

BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION AL108460

VERSION AL108460.1 GI:5628764

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## Location/Qualifiers

## 1..1101

## /organism="Drosophila melanogaster"

## /plasmid="pBelobAC11"

## /db\_xref="taxon:7227"

## /clone\_lib="DrosBAC"

## /clone="BACN37L08"

## /note="end : SP6"

## BASE COUNT 254 a 176 c 160 g 152 t 359 others

## ORIGIN

## Query Match 11.8%; Score 48.2; DB 12; Length 1101;

## Best Local Similarity 13.6%; Pred. No. 35;

## Matches 45; Conservative 169; Mismatches 114; Indels 3; Gaps 1;

## Qy 14 tctccgagagctctcgttcggtattccggtgagctccgatacaggtcgcgatccgt 73

## Db 1072 TBSSBSTSTSTBTBSSSSSSBBSBSSSSSSBSSSSSSSSSSSSSTSSSSSTSB 1013

## Qy 74 atgcacccgagatgcgttcacacctccc---ggcgatgccctgtgacctggcggtcg 130

## Db 1012 BTSTSSSSSTTTTTTTTTTTTBTBSSBSSBSSSSSTSSSBSTSKBTSBSSB 953

## Qy 131 gccgcgagctctgtagcaggggtcaacagcccgagcgagcgagcgatgtgcacatcg 190

## Db 952 SBSSSSSSSBSTSTBTBSSBSSSSSSSSSSSSSSSSSSSSSSSSSTBTBT 893

## Qy 191 gccgcagcagcccgagggcctcgagatgtccacatccggtccaggtcgcgcgagcc 250

## Db 892 TSSSSTBTBTBSSSSSSSSSSSSSSBSSSTTSSBSSBSSSSBCTSKCSTBT 833

## Qy 251 gtgcgctgttcggcggggagcgagcgctggtggttcttcgacgagcgagcgagc 310

## Db 832 STBMSKBSSTSTSTBTSBSCGCGSTSGSCSBCGKSTSGSTSSGTTTCGBCS 773

## Qy 311 tcgtgcgctcggccagagcagcagcagcgtggg 341

## Db 772 GCSCCCYCCCT 742

Search completed: July 18, 2002, 10:12:46

Job time: 9966 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 11:22:50 ; Search time 130.38 Seconds  
(without alignments)  
115.010 Million cell updates/sec

Title: US-09-749-185-3

Perfect score: 704

Sequence: 1 MSFLVSEELSFRIPELVRYE.....FDGNLEDAIGRLAEEQNAG 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	704	100.0	135	21	AA144649
2	606	86.1	135	21	AA144652
3	556	79.0	135	21	AA144651
4	464	65.9	135	21	AA144650
5	84	11.9	321	22	ABG24744
6	83.5	11.9	456	21	AAG12893
7	83.5	11.9	456	21	AAG27402
8	81	11.5	552	18	AA14287
9	81	11.5	552	21	AA170071
10	81	11.5	552	21	AA169622
11	81	11.5	798	19	AA185025
					Streptomyces griseus
					Streptomyces netro
					Streptomyces golde
					Streptomyces albus
					Novel human diagno
					Arabidopsis thalia
					Arabidopsis thalia
					Human deleted in p
					Human tumour suppr
					Human Smad4. Homo
					Smad4-green floure

12	81	11.5	806	19	AA185013
13	75.5	10.7	1046	17	AA102156
14	75.5	10.7	1046	21	AA152304
15	74	10.5	400	22	AA156775
16	73	10.4	674	22	AB164343
17	72	10.2	1248	20	AA143569
18	72	10.2	1248	20	AA123960
19	72	10.2	1248	21	AA167527
20	72	10.2	3596	21	AA187407
21	72	10.2	3647	11	AA105041
22	71.5	10.2	212	21	AA106133
23	71.5	10.2	234	20	AA106221
24	71.5	10.2	971	22	AA1050418
25	71	10.1	837	21	AA112381
26	71	10.1	1248	19	AA169594
27	70.5	10.0	787	22	AB171973
28	70	9.9	448	22	AB153026
29	69.5	9.9	252	21	AA124694
30	69.5	9.9	259	22	AA159590
31	69.5	9.9	317	21	AA124693
32	69.5	9.9	364	21	AA124692
33	69.5	9.9	1298	22	AB161948
34	69	9.8	752	22	AA149339
35	68.5	9.7	361	21	AA175618
36	68	9.7	143	21	AA181424
37	68	9.7	323	22	AA161464
38	68	9.7	401	22	AA135669
39	68	9.7	410	22	AB161295
40	68	9.7	4274	22	AB100972
41	68	9.7	4386	22	AB107375
42	68	9.7	4397	22	AB121944
43	67.5	9.6	108	21	AA102229
44	67.5	9.6	154	22	AB152719
45	67.5	9.6	306	21	AA115194

#### ALIGNMENTS

#### RESULT 1

AA144649  
ID AA144649 standard; Protein; 135 AA.

XX AC AA144649;

XX DT 18-APR-2000 (first entry)

XX DE Streptomyces griseus SsgA protein.

XX SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.

XX OS Streptomyces griseus.

XX PN WO200000613-A1.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99WO-NL00395.

XX PR 26-JUN-1998; 98EP-0202148.

XX PA (UYLE-) RIJKSUNIV LEIDEN.

XX PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.

XX PI Van Wezel GP, Kraal B, Luiten RCM;

XX DR WPI; 2000-147269/13.

XX DR N-PSDB; AA149727, AA149728.

XX





Filamentous bacteria can be transformed with *ssrA* gene-containing plasmid to enhance the production of secondary metabolites such as antibiotics, antitumour agents, immunosuppressive agents, hypcholesterolaemic agents, enzyme inhibitors, antimigraine agents, herbicides, antiparasitic agents, ruminant growth promoters, bioinsecticides, receptor agonists and antagonists and biomass.

Query Match	79.0%;	Score 556;	DB 21;	Length 135;
Best Local Similarity	78.5%;	Pred. NO. 3.5e-55;		
Matches 106;	Conservative 12;	Mismatches 17;	Indels 0;	Gaps 0;
Qy	1	MSFLVSEELSFRIPELVRYEVGGDPYAIRMTFHPVGGAPVTWAFGRRELLDGLNPSGDDG 60		
Db	1	msflvseelsfrfipvelryevggdpayairmtfhpvggapvtwafgrrelldglngprpcgdd 60		
Qy	61	VHIGTPEGIGDVHIRLQVAGDRAALFRAGCTAPLVAFLDRTDKLVLPGVQHTLTGDFDGNL 120		
Db	61	vhiaapdrfgevlrlqvgdqamfrvgtaplvafldrtkivlpgqersladfdall 120		
Qy	121	EDALGRILAEQNAQ 135		
Db	121	dealdrillaeqnaq 135		

RESULT	4	
AA44650		
ID	AA44650	standard; Protein; 135 AA.
XX		
XX	AA44650;	
XX		
DT	18-APR-2000	(first entry)
XX		
XX	Streptomyces albus	G SsgA protein.
XX		
XX	SsgA;	liquid culture; filamentous bacteria; secondary metabolite;
KW	mycelium;	antibiotic; antitumour agent; immunosuppressive agent;
KW	hypocholesterolaemic agent;	enzyme inhibitor; antimigraine agent;
KW	hericide; antiparasitic agent;	ruminant growth promoter;
KW	biolinsecticide; receptor agonist;	antagonist; biomass.
XX		
XX	Streptomyces albus	G.
XX		
XX	WO200000613-A1.	
PN		
XX		
PD	06-JAN-2000.	
XX		
XX	25-JUN-1999;	99WO-NL00395.
XX		
XX	26-JUN-1998;	98EP-0202148.
PR		
XX		
PA	(UYLE-) RIJKSUNIV LEIDEN.	
PA	(NEWE-) NEDERLANDSE ORG	WETENSCHAPPELIJK ONDERZO.
XX		
PI	Van Wezel GP, Kraal B, Luiten RGM;	
XX		
DR	WPI; 2000-147269/13.	
DR	N-PSDB; AAZ49729.	
XX		
XX	Reducing branching and enhancing fragmentation in filamentous	
PT	microorganisms used to improve their liquid culturing properties	-
XX		
PS	Disclosure; Fig 5; 60pp; English.	
XX		
XX	The present sequence is S. albus G SsgA protein. SsgA reduces branchi	
CC	and fragment septation and enhances fragmentation of mycelium in liqu	
CC	culture resulting in lower viscosity of culture broths. Filamentous	
CC	bacteria can be transformed with ssgA gene-containing plasmid to enhan	
CC	the production of secondary metabolites such as, antibiotics, antitum	
CC	agents, immunosuppressive agents, hypocholesterolaemic agents, enzyme	
CC	inhibitors, antimigraine agents, herbicides, antiparasitic agents,	

CC	ruminant growth-promoters, bioinsecticides, receptor agonists and
CC	antagonists and biomass.
XX	
SQ	Sequence 135 AA;
	Query Match 65.9%; Score 464; DB 21; Length 135;
	Best Local Similarity 65.9%; Pred. No. 9.7e-45;
	Matches 89; Conservative 18; Mismatches 28; Indels 0; Gaps 0;
QY	1 MSFLVSEELGFRIPVELRYEVDGYPYAIRMTFHLPGDAPVTWAFGRELLDGLNLSPSGDGD 60
DB	1 mslvseelafrripvelryetvdpvavrllfhlpgdpavtwfgrllvegyidaagdg 60
QY	.61 VHIGTPEGLGDVHRIQVGADRALFRAGTAPILVAFLDRTDKLIVPLGQEHLLGDFDGNL 120
DB	61 vrvcpvgqtarevhitlqvgsqalfrvgkapllafldrtddgslgserahadfgshl 120
QY	121 EDALGRILAEQNAG 135
DB	121 ddalnrslaeeqsag 135
RESULT	5
ABG24744	
ID	ABG24744 standard; Protein; 321 AA.
XX	
AC	ABG24744;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #24735.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WC200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	N-PSDB; AAS88931.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	Claim 20; SEQ ID No 55103; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AEG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 321 AA;

Query Match 11.9%; Score 84; DB 22; Length 321;  
Best Local Similarity 27.7%; Pred. No. 0.39;  
Matches 44; Conservative 21; Mismatches 52; Indels 42; Gaps 9;

Qy 1 MSFLVSELSFRIPVELRYEGDYPVIRMTFHLPGDAPVTWAFG----- 44  
Db 28 ltfmklrdirdpmlldevenelki-----gcapitwpgcgklfkgvyhlykdet 80  
Qy 45 ---REL-LLDGLNSPSGSDVHIGTEPEGLGDVHIRLQVCA----DRALFRAGTAPLVA 96  
Db 81 yllgevrivkglnnp--dldaavgedlaqlrd-elelvkgasnefkdflageitpvf 137  
Qy 97 FLDRDVKLVPLGQHTLGDF---DGNLEDAL-GRILAE 130  
Db 138 f--gtalgnfgvdmldgmfnagdrshveaypgdliq 173

RESULT 6  
AAG12893  
ID AAG12893 standard; Protein: 456 AA.

XX AAG12893;  
XX 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 12181.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.  
PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127452.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 06-MAY-1999; 99US-0132487.

XX 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.



PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142354.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0156559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.

```
(in vivo or in vitro) proliferative diseases, especially pancreatic carcinoma, bile duct, bladder or colorectal cancer, Crohn's disease, colitis-associated neoplasia or chronic ulcerative colitis. These conditions, where associated with a homozygous deletion, can be treated by administering an agent that: (a) modulates DPC4 expression, specifically a sense DPC4 sequence (particularly in the form of a vector, i.e. by gene therapy), but also an antisense sequence where DPC4 protein is over expressed or (b) mimics the activity of DPC4. DPC4 nucleic acid is also used as hybridisation probes for detecting presence/absence of human chromosome 18q21.1 fragments. When a homozygous deletion is detected in this region, an agent can be administered that accumulates within, or kills, only cells which contain such a deletion. This agent exploits the absence of an enzyme (or other protein) encoded by a neighbouring gene and lost by the CC deletion, i.e. it has a highly selective action.
```

```
XX  
SQ Sequence      552 AA;  
  
Query Match          11.5%; Score 81; DB 18; Length 552;  
Best Local Similarity 25.7%; Pred. No. 1.7;  
Matches    38; Conservative   16; Mismatches    28; Indels     66; Gaps           9;
```

```
QY  5 VSELSFRIPV-----ELRYEVGDPPIAIRMTFHLPGDAPVTWAFGRELLL 49  
| |||| : | : : || : || : || : || : || : || : || :  
Db  304 vlnelaifqpishnpapeywcslayfemdvqvge-----tfkvpsscpi-----vtv 350
```

```
QY  50 DGLNSPPGDGG-----VHTG---PTEPEGLGVHRLQVGADRAL 86  
|| || || || || || || || || || || || || || || || :  
Db  351 dgyvdpsg-gdrfcqlgsnvhrteatarlhlgkgvleckegdwwrcrl--sdhav 407
```

```
QY .87 F-----RAGTAPLVAFLDRTDKLPV 106  
| || || || || || || || || || || || || || || || :  
Db  408 fvqsyyldreagrap----gdavkhkiyp 431
```

```
RESULT       9  
AAY70071  
ID AAY70071 standard; Protein; 552 AA.  
AC AAY70071;  
XX  
XX  
DT DT  
XX Human tumour suppressor protein, Pdc4.  
DE DE  
XX  
KW Call proliferative disorder; nuclear localisation factor; neoplasm;  
KW Dpc4; Deleted in Pancreatic Carcinoma; locus 4; Smad-binding element;  
KW SBE; tumor suppressor; transforming growth factor-beta; TGF beta;  
KW anti-cancer drug; treatment; gene therapy; human.  
XX  
OS Homo sapiens.  
XX  
XX W0200009526-A2.  
PN  
XX  
PD 24-FEB-2000.  
XX  
XX 13-AUG-1999; 99WO-US18540.  
PF  
XX 14-AUG-1998; 98US-0096628.  
PR  
XX (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
PA Kern SE, Dai JL;  
PI  
XX WP1::2000-224266/19.  
XX N-PSDB; AA251013.  
PT Treatment of a cell proliferative disorder by administration of tumor  
PT suppressor polypeptide Dpc4 (Smad4) coupled to a nuclear localization  
XX factor -  
XX Claim 1; Page 68; 68pp; English.
```

XX	The patent discloses a method of treating cell proliferative disorders,
CC	using a chimeric Dpc4 (Deleted in Pancreatic Carcinoma, locus 4)
CC	polypeptide coupled to a nuclear localisation factor. Upon localisation
CC	to the nucleus and binding to Smad-binding element (SBE), Dpc4 shows
CC	tumour suppressor action. This method can also be used for identifying
CC	transforming growth factor-beta (TGF beta) inducible genes, modulators
CC	of Dpc4 nuclear localisation and in screening for anti-cancer drugs.
CC	Dpc4 can be used in the treatment of neoplasms and in gene therapy.
CC	The present sequence is that of human Dpc4 tumour suppressor protein.
CC	Note: It is indicated in the claims section that the present peptide
CC	sequence is, sequence Id No: 1 and elsewhere in the specification as
CC	seq id No: 2.
XX	Sequence 552 AA;
SQ	

Query Match	11.5%;	Score 81;	DB 21;	Length 552;
Best Local Similarity	25.7%;	Pred. NO. 1.7;		
Matches 38;	Conservative	16;	Mismatches 28;	Indels 66;
				Gaps 9;

[illegible]

RESULT	10
AA	AY69622
ID	AY69622 standard; Protein; 552 AA.
XX	
AC	AY69622;
XX	
DT	19-APR-2000 (first entry)
XX	
DE	Human Smad4.
XX	
KW	Smad4; MADH4; DPC4; TGF-beta signalling pathway; transcription factor;
KW	expression inhibition; antisense therapy; tumour formation;
KW	inflammation.

CC superfamily signal transduction. On ligand binding, TGF-beta superfamily  
CC proteins (such as bone morphogenetic protein (BMP), activin and  
CC TGF-betas themselves) phosphorylate Smad proteins, which then homo- or  
CC heterodimerise and translocate to the nucleus to activate target gene  
CC transcription. Smad4 (also known as MADH4 and DPC4) is a shared  
CC heterodimerisation partner for the pathway recognised members of the  
CC Smad family (Smad1-3, 5 and MADH6) and is known as the common mediator.  
CC The N-terminus of Smad4 promotes the binding of the Smad complex to DNA,  
CC and the C-terminus provides an activation signal required for the  
CC complex to stimulate transcription. The invention relates to antisense  
CC oligonucleotides targeted to the human Smad4 gene, which inhibit its  
CC expression. A series of oligonucleotides (AA259749-259788) were designed  
CC to target different regions of the human Smad4 RNA, and were analysed  
CC for their effect on Smad4 mRNA levels by quantitative real-time PCR. The  
CC oligonucleotides of the invention are useful for diagnosis, prevention  
CC and treatment of conditions associated with Smad4 expression, such as  
CC tumour formation, inflammation and certain infections.

```
Query Match      11.5%; Score 81; DB 21; Length 552;
Best Local Similarity 25.7%; Pred. No. 1.7;
Matches 38; Conservative 16; Mismatches 28; Indels 66; Gaps 9;
```

Qy	5	VSELSRPV-----ELRYEVGGPYAIRMTFHLPGDAPVTWATGRELLL 49
Dd	304	vhnelaqpslnhpapeycwclayfemdvge-----tfkypscpi-----vfv 350
		:   :   :   :   :   :   :   :   :   :   :
Qy	50	DGLNPSGSDG-----VHIG----PTEPEGIQGVHRIQLQVGADRAL 86
		:   :   :   :   :   :   :   :   :   :   :
Dd	351	dgyvdpsg-gdrfcqlsnvhrtatearhlhgkgvleckegdwvrcl--sdhav 407
Qy	87	F-----RAGTAPLVAFLDRTOKLVP 106
Dd	408	fvgvyidreagrap---gdavhkivp 431

RESULT	11
AAW85025	AAW85025 standard; Protein; 798 AA.
XX	XX
XX	AAW85025;
XX	XX
XX	08-FEB-1999 (first entry)
DT	XX
XX	Smad4-green fluorescent protein fusion product.
DE	XX
XX	Human; Smad4 gene; fusion protein; green fluorescent protein; GFP;
XX	Intracellular signalling; Chimera.
KW	XX
KW	Chimeric - Aequorea victoria.
OS	Chimeric - Homo sapiens.
OS	XX
XX	WO9845704-A2.
PN	XX
XX	15-OCT-1998.
XX	XX
PD	XX
XX	07-APR-1998; 98WO-DK00145.
PF	XX
XX	XX
XX	07-APR-1997; 97DK-0000392.
PR	XX
XX	(NOVO ) NOVO-NORDISK AS.
PA	XX
XX	Kasper A, Petersen Bjorn S, Scudder K, Thastrup O;
PI	Tullin S;
PI	XX
XX	WPI; 1998-594491/50.
DR	N-PSDB; AAW71040.
DR	XX
XX	Determining effect on signalling pathways in live cells from
PT	redistribution of luminophores - specifically fusions of green
PT	fluorescent protein with a signalling component, and new apparatus
PT	XX



XX Periplasmic chitodextrinase (AAW02156), periplasmic  
CC Beta-N-acetylglucosaminidase (AAW02157) and aryl  
CC Beta-N-acetylglucosaminidase (AAW02158) can be used to produce chitin  
CC oligosaccharides with the structure (GlcNAc)<sub>n</sub> where n is 2 or  
CC higher, by contacting them with soluble chitin. The enzymes are  
CC encoded by the genes endI, exoI and exoII respectively. They are  
CC all genes involved in the catabolic pathway of chitin.  
XX  
SQ Sequence 1046 AA;

Query Match 10.7%; Score 75.5; DB 17; Length 1046;  
Best Local Similarity 30.5%; Pred. No. 17;  
Matches 25; Conservative 8; Mismatches 30; Indels 19; Gaps 3;

QY 9 LSPRIPVELRYEVDPPAIRMTFHLPGDAP-----VTWAFGRRL-----LLDGL 52  
Db 920 vafslpawktlpagdtvdyldmvyvlpisgpanysvningvdyafkfeqdpdladlssgn 979  
QY 53 NSPSGDDGVHIGPTEPEGLGDV 74  
Db 980 ggggtggdgtgggttep---gdv 998

RESULT 14  
AAV52304  
ID AAV52304 standard; Protein; 1046 AA.  
AC AAV52304;  
XX  
DT 09-FEB-2000 (first entry)  
DE Vibrio furnissii periplasmic chitodextrinase.  
XX  
KW Chitodextrinase; periplasmic; endI; Endo-I; chitin; chitosan; endoenzyme;  
KW beta 1-4 N-acetylglucosamine; GlcNAc; degradation; catabolism;  
KW oligosaccharide; agriculture; medicine; nitrogen fixing nodules;  
KW disease resistance; induction; fungicide; antimetastatic;  
KW Lewis lung carcinoma; immune system; macrophage; activation; production;  
KW recombinant protein.  
XX  
OS Vibrio furnissii.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..30  
FT /note= "Periplasmic localisation signal peptide"  
FT Protein 31..1046  
FT /note= "Mature periplasmic chitodextrinase"  
XX  
XX US5985644-A.  
XX  
XX 16-NOV-1999.  
XX  
XX 13-FEB-1996; 96US-0600452.  
XX  
XX 13-FEB-1995; 95US-0386727.  
XX  
XX (UJJO ) UNIV JOHNS HOPKINS.  
XX  
XX Bassler B, Chitlaru E, Yu C, Roseman S, Keyhani NO;  
XX WPI; 2000-022280/02.  
XX N-PSDB; AA238240.  
XX  
XX DNA encoding periplasmic chitodextrinase endoenzyme -  
XX  
XX Claim 1; Columns 33-40; 37pp; English.  
XX  
XX This sequence represents Vibrio furnissii periplasmic chitodextrinase  
CC (Endo-I) which is encoded by the endI gene. Chitin is a homopolymer of  
CC beta 1-4 N-acetylglucosamine (GlcNAc). Chitin degradation by V. furnissii  
CC involves several signal transducing systems and a multitude of proteins,  
XX

CC in contrast to other organisms, which only require two enzymes to degrade  
CC chitin to GlcNAc. Endo-I is an endoenzyme, cleaving internal glycosidic  
CC bonds in soluble chitin-derived oligosaccharides to produce GlcNAc dimers  
CC and/or trimers. Chitin oligosaccharides have been recently shown to be  
CC physiologically active and are useful in agriculture and medicine.  
CC Derivatised oligosaccharides are generated by Rhizobium species as  
CC signals for the formation of nitrogen fixing nodules by leguminous plants  
CC and also induce disease resistance in certain plants. They also inhibit  
CC the growth of several fungal pathogens. The GlcNAc hexamer is a potent  
CC antimetastatic agent against Lewis lung carcinoma, and GlcNAc polymers of  
CC varying lengths activate macrophages and the immune system. Prior art  
CC methods of production of such oligosaccharides are prohibitively  
CC expensive and there are limitations in resolving mixtures of these  
CC compounds. Use of recombinant V. furnissii chitin catabolic enzymes  
CC may help to reduce or eliminate these problems.  
XX  
SQ Sequence 1046 AA;

Query Match 10.7%; Score 75.5; DB 21; Length 1046;  
Best Local Similarity 30.5%; Pred. No. 17;  
Matches 25; Conservative 8; Mismatches 30; Indels 19; Gaps 3;

QY 9 LSPRIPVELRYEVDPPAIRMTFHLPGDAP-----VTWAFGRRL-----LLDGL 52  
Db 920 vafslpawktlpagdtvdyldmvyvlpisgpanysvningvdyafkfeqdpdladlssgn 979  
QY 53 NSPSGDDGVHIGPTEPEGLGDV 74  
Db 980 ggggtggdgtgggttep---gdv 998

RESULT 15  
AAU56775  
ID AAU56775 standard; Protein; 400 AA.  
XX  
AC AAU56775;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #17671.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
XX WO200181581-A2.  
XX  
XX 01-NOV-2001.  
XX  
XX 20-APR-2001; 2001WO-US12865.  
XX  
XX 21-APR-2000; 2000US-199047P.  
XX  
XX 02-JUN-2000; 2000US-208841P.  
XX  
XX 07-JUL-2000; 2000US-216747P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
XX WPI; 2001-616774/71.  
XX N-PSDB; AAS59578.  
XX  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
XX vaccinating against and diagnosing infections, especially useful for  
XX treating acne vulgaris -  
XX  
XX Claim 6; SEQ ID NO 17970; 1069pp; English.  
XX



Search completed: July 18, 2002, 14:15:01  
Job time: 10331 sec

**THIS PAGE BLANK (USPTO)**



```
Db 351 DGYVDPG-GDRFCGLQLSNVHRTAIERARLHIGKVQLECKGEGVWVRCL--SDHAV 407
Qy 87 F-----RAGTAPLVAFLDRTDKLVP 106
Db 408 FVQSYLDREAGRAP-----GDAVHKIYP 431

RESULT 2
US-08-915-214-2
; Sequence 2, Application US/08915214
; Patent No. 5814457
; GENERAL INFORMATION:
; APPLICANT: Kern, Scott E.
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, DPC4
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,214
; FILING DATE: 20-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/588,821
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/079001
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-214-2

Query Match 11.5%; Score 81; DB 2; Length 552;
Best Local Similarity 25.7%; Pred. No. 0.27; Indels 66; Gaps 9;
Matches 38; Conservative 16; Mismatches 28;

Qy 5 VSEELSFRIYV-----ELRYEYGDYAIRMTFHLPGDAPVTWAFGRELLL 49
Db 304 VHNELAFQPPISNHPAPEYWCYSIAVFENDVQVGE-----TFKVPSSCPI-----VTV 350
Qy 50 DGLNSPSCGDG-----VHIG---PTEPEGLGDVHIRLQVGADRAL 86
Db 351 DGYVDPG-GDRFCGLQLSNVHRTAIERARLHIGKVQLECKGEGVWVRCL--SDHAV 407
Qy 87 F-----RAGTAPLVAFLDRTDKLVP 106
Db 408 FVQSYLDREAGRAP-----GDAVHKIYP 431

RESULT 3
US-09-005-532-2
; Sequence 2, Application US/09005532
; Patent No. 5955292
```

```
; GENERAL INFORMATION:
; APPLICANT: Kern, Scott E.
; APPLICANT: Hahn, Stephan A.
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, DPC4
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,532
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/588,821
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/079001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-005-532-2

Query Match 11.5%; Score 81; DB 2; Length 552;
Best Local Similarity 25.7%; Pred. No. 0.27; Indels 66; Gaps 9;
Matches 38; Conservative 16; Mismatches 28;

Qy 5 VSEELSFRIYV-----ELRYEYGDYAIRMTFHLPGDAPVTWAFGRELLL 49
Db 304 VHNELAFQPPISNHPAPEYWCYSIAVFENDVQVGE-----TFKVPSSCPI-----VTV 350
Qy 50 DGLNSPSCGDG-----VHIG---PTEPEGLGDVHIRLQVGADRAL 86
Db 351 DGYVDPG-GDRFCGLQLSNVHRTAIERARLHIGKVQLECKGEGVWVRCL--SDHAV 407
Qy 87 F-----RAGTAPLVAFLDRTDKLVP 106
Db 408 FVQSYLDREAGRAP-----GDAVHKIYP 431

RESULT 4
US-08-701-582D-14
; Sequence 14, Application US/08701582D
; Patent No. 6017755
; GENERAL INFORMATION:
; APPLICANT: WRANA, Jeffrey
; APPLICANT: ATTISANO, Lilliana
; APPLICANT: SCHERER, Stephen W.
; TITLE OF INVENTION: MADR2 TUMOUR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
```

Query Match	11.5%	Score 81;	DB 4;	Length 552;
Best Local Similarity	25.7%	Pred. No. 0.27;		
Matches 38;	Conservative 16;	Mismatches 28;	Indels	



EARLIER APPLICATION NUMBER: DE 198 46 493.2  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 93  
LENGTH: 3072  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-93

Query Match 10.4%; Score 73; DB 4; Length 3072;  
Best Local Similarity 31.4%; Pred. No. 22;  
Matches 33; Conservative 7; Mismatches 59; Indels 6; Gaps 3;  
Qy 1 MSFLVSELSRIPVELRYEGDPAIRMTFHLPGDAPVTWAFGRELLDGLNSPSGDG 60  
Db 869 MOVLGRIESSLGRITLTLFHEFTPLAQLASHLSSGAATAAALERGLTRDPGSS 928  
Qy 61 VHIG-PTEP-----EGLGDVHRLQVGADRALFRAG-TAPLVAFLD 99  
Db 929 PRVATPEEPFALTEGORAMWLECKQSDAGALYNLGRTVRLGAGVD 973

RESULT 10  
US-09-413-814-80  
Sequence 80, Application US/09413814  
Patent No. 6225064  
GENERAL INFORMATION:  
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
APPLICANT: Bristol-Myers Squibb, Co.  
APPLICANT: Beyer, Stefan  
APPLICANT: Bloecker, Helmut  
APPLICANT: Brandt, Petra  
APPLICANT: Cino, Paul M  
APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hofle, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
TITLE OF INVENTION: heteropolypeptide compounds  
FILE REFERENCE: PCT/US 99/23535  
CURRENT APPLICATION NUMBER: US/09/413,814  
CURRENT FILING DATE: 1999-10-07  
EARLIER APPLICATION NUMBER: DE 198 46 493.2  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 80  
LENGTH: 3079  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-80

Query Match 10.4%; Score 73; DB 4; Length 3079;  
Best Local Similarity 31.4%; Pred. No. 22;  
Matches 33; Conservative 7; Mismatches 59; Indels 6; Gaps 3;  
Qy 1 MSFLVSELSRIPVELRYEGDPAIRMTFHLPGDAPVTWAFGRELLDGLNSPSGDG 60  
Db 875 MOVLGRIESSLGRITLTLFHEFTPLAQLASHLSSGAATAAALERGLTRDPGSS 934  
Qy 61 VHIG-PTEP-----EGLGDVHRLQVGADRALFRAG-TAPLVAFLD 99  
Db 935 PRVATPEEPFALTEGORAMWLECKQSDAGALYNLGRTVRLGAGVD 979

RESULT 11  
US-08-348-353-17  
Sequence 17, Application US/08348353  
Patent No. 5932217

GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine  
APPLICANT: Masure, Robert  
TITLE OF INVENTION: Antibody Recognizing Endothelial Cell  
TITLE OF INVENTION: Ligand for Leukocyte CR3  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/348,353  
FILING DATE: 30-NOV-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson, David  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-097CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1248 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-348-353-17

Query Match 10.2%; Score 72; DB 2; Length 1248;  
Best Local Similarity 27.6%; Pred. No. 8.7;  
Matches 29; Conservative 11; Mismatches 43; Indels 22; Gaps 4;  
Qy 47 LLLDG-----LNSPSGDGDVH-----IGPTEPEGLGDVHRLQVGADRALF 87  
Db 53 LALDGGAGVALQSAKASGTLHVQGGHLDLGLTAAVAGVDVNGTGDVRYAKLVSDAGADL 112  
Qy 88 RAGTAPLVAFLDRTDKLVPLQGEH-TLGDFF--DGNLEDAALGRILA 129  
Db 113 QAGSMTLGIIVDTTGDLOARAQKLELGSVKSDGGLQAAAGGALS 157

RESULT 12  
US-08-465-965-17  
Sequence 17, Application US/08465965  
Patent No. 5968512

GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine  
APPLICANT: Masure, Robert  
TITLE OF INVENTION: Antibody Recognizing Endothelial Cell  
TITLE OF INVENTION: Ligand for Leukocyte CR3  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

REPLY 13  
US-08-465-966-17  
Sequence 17, Application US/08465966  
Patent No. 6015560  
GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine  
APPLICANT: Mesure, Robert  
TITLE OF INVENTION: Antibody Recognizing Endothelial Cell  
Ligand for Leukocyte CR3  
TITLE OF INVENTION: Ligand for Leukocyte CR3  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465-966

```

RESULT 14
US-08-857-076-114
; Sequence 114, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowsek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-857-076-114

```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 11:38:30 ; Search time 73.98 Seconds  
(without alignments)  
175.345 Million cell updates/sec

Title: US-09-749-185-3

Perfect score: 704

Sequence: 1 MSFLVSEELSFRIPELVRYE.....FDGNLEALGRILAEQNAQ 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	546	77.6	136	2	T37179	probable regulator
2	219	71.1	142	2	T36147	probable regulator
3	190	27.0	138	2	T35247	probable regulator
4	140	19.9	142	2	T35319	probable regulator
5	84.5	12.0	548	2	E94182	type II secretion
6	84	11.9	261	2	T19428	hypothetical prote
7	83.5	11.9	259	2	A22612	carbonate dehydrat
8	81	11.5	324	2	A69648	2-keto-3-deoxygluc
9	81	11.5	552	2	S71811	probable transcrip
10	79.5	11.3	408	2	A32377	conserved hypothet
11	78.5	11.2	260	2	I52551	carbonate dehydrat
12	78.5	11.2	1101	2	S15271	endoglucanase C (E
13	78	11.1	529	2	E91295	peptide chain rele
14	78	11.1	529	2	I59305	translation releas
15	78	11.1	529	2	G86136	peptide chain rele
16	78	11.1	690	2	D75487	v-type ATP synthas
17	77.5	11.0	260	2	A43641	carbonate dehydrat
18	77.5	11.0	913	2	T35718	hypothetical prote
19	77	10.9	795	1	A70322	endopeptidase La (
20	77	10.9	819	2	T45690	receptor-like prot
21	77	10.9	827	1	A36895	endopeptidase La (
22	76.5	10.9	757	2	G97472	penicillin-binding
23	76.5	10.9	757	2	AC2691	penicillin binding
24	75.5	10.7	902	2	G83635	probable ClpA/B-ty
25	75.5	10.7	1046	2	T30199	chitinase (EC 3.2
26	75	10.7	760	2	E83610	conserved hypothet
27	73.5	10.4	265	2	C70963	hypothetical prote
28	73.5	10.4	346	1	H69789	probable alcohol d
29	73.5	10.4	484	2	S68694	hexokinase (EC 2.7

translation releas  
probable acetylorn  
hypothetical cytos  
peptide chain rele  
probable 9-cis-epo  
tRNA-intron endonu  
2-oxoacid--ferredo  
endopeptidase La-1  
peroxisome biogene  
probable ATP-depen  
filamentous hemagg  
aminonucleoside an  
clp ATPase [import  
CDA peptide synthet  
hypothetical protei  
DNA-Binding Protei

ALIGNMENTS

RESULT 1

T37179

probable regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T37179

R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: Z21598

A:Accession: T37179

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-136 <SEE>

A:Cross-references: EMBL:AL096823; PIDN:CAB46964.1; GSPDB:GN00070; SCOEDB:SCQ11.09

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: ssgA; SCOEDB:SCQ11.09

Query Match

Best Local Similarity 77.6%; Score 546; DB 2; Length 136;

Matches 105; Conservative 12; Mismatches 18; Indels 0; Gaps 0;

QY 1 MSFLVSEELSFRIPELVRYEVDPAIRMTFHLPGDAPVTWAFGRLLDGLNPSGSDG 60

DB 2 MSFLVSEELSFRIPELVRYETRDPTAVRLTFHLPGDAPVTWAFGRLLDGLNPSGSDG 61

QY 61 VHIGTEPEGLGDVHRLQVGADRALFRAGTAPLVAFLDRTDKLVLPGQEHITLGDGDL 120

DB 62 VRIAPVEPEPLAEVLRLQVGSQDALFRSSAAPLVAFLDRTDKLVLPGQEGALADFSDL 121

QY 121 EDALGRILAEQNAQ 135

DB 122 DEALDRILAEQNAQ 136

RESULT 2

T36147

probable regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000

C:Accession: T36147

R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: Z21598

A:Accession: T36147

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-142 <SEE>

A:Cross-references: EMBL:AL096852; PIDN:CAB51005.1; GSPDB:GN00070; SCOEDB:SCE19A.24

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC19A.24  
C:Superfamily: Streptomyces coelicolor probable regulator SCE19A.24

Query Match 31.18; Score 219; DB 2; Length 142;  
Best Local Similarity 37.7%; Pred. No. 1.6e-14;  
Matches 49; Conservative 22; Mismatches 53; Indels 6; Gaps 2;  
Qy 1 MSFLVSELSRIPVELRYEGDPAVIRMTFLPGDAPVTWAFGRELLDGLNSPSPGD 60  
Db 17 LRLVLSPEGPVPAFLGYHDDPYAVHITHDSGHPVHVTFAFDLLVEGVFRPSGHG 76  
Qy 61 VHIGTEPEGLGVHRLQVGADRALFRAGTAPLVAFLDRTDKLVLQGEHT-LGDFDGN 119  
Db 77 VRVWPSKTEGRSVLVALLSSPDGALLEAPTPQYSAWLERTLRAVPPCTEGAQLG----- 131  
Qy 120 LEDALGRILA 129  
Db 132 IDDLGLAELLA 141

RESULT 3  
T35247  
Probable regulator - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 20-Jun-2000  
C:Accession: T35247  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, April 1999  
A:Reference number: Z21573  
A:Accession: T35247  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-138 <OLI>  
A:Cross-references: EMBL:AL049587; PIDN:CAB40672.1; GSPDB:GN00070; SCOEDB:SC5F2A.05c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC5F2A.05c  
C:Superfamily: Streptomyces coelicolor probable regulator SCE19A.24

Query Match 27.08; Score 190; DB 2; Length 138;  
Best Local Similarity 31.3%; Pred. No. 1.1e-11;  
Matches 46; Conservative 19; Mismatches 46; Indels 36; Gaps 4;

Qy 1 MSFLVSELSR-----IPVELRYEGDPAVIRMTFLPG---DAPVTWAFGREL 47  
Db 1 MSTVIGSVEARLVAAAPRPSIPATLHYRADPFAVRMTFPAPATLEGVEVCWTFSGREL 60  
Qy 48 LLDGLNSPSPGDGVHIGP-----TEPEGLGVHRLQVGADRALFRAGTAPLVAFL 97  
Db 61 LIAGMQEPNGHGVVRVPAVYDRVTLEFHAPEGTAIVHR-----SGELRRF 107  
Qy 98 LDRTDKLVLQGEHTLGDFDGNLEDAL 124  
Db 108 LQAAGELVPVGLLEHLQDLQDLHDLAELM 134

RESULT 4  
T35319  
Probable regulator - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 20-Jun-2000  
C:Accession: T35319  
R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z21575  
A:Accession: T35319  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-142 <OLI>  
A:Cross-references: EMBL:AL049863; PIDN:CAB42928.1; GSPDB:GN00070; SCOEDB:SC5H1.03  
A:Experimental source: strain A3(2)

C:Genetics:  
A:Gene: SCOEDB:SC5H1.03  
C:Superfamily: Streptomyces coelicolor probable regulator SCE19A.24

Query Match 19.9%; Score 140; DB 2; Length 142;  
Best Local Similarity 31.0%; Pred. No. 1e-06;  
Matches 40; Conservative 20; Mismatches 51; Indels 18; Gaps 5;  
Qy 12 RIPV--ELRYEVGDPYAIRMTFLPGDAPVTWAFGRELLDGLNSPSPGDGVHIGPTEPE 69  
Db 21 RFPVLAHLAYDAADPFALTVVFSHDGRVRLARMTLDREMAEGLTRPVGVGVDRLRP-ESR 79  
Qy 70 GLGVHRLQVGAD-----RALFRAGTAPLVAFLDRTDKLVLQGEHTLGDFDGNLE 121  
Db 80 GMWD-ELRIELLDGRADGERHRAVVFVMAAAVEAFLRETHAVVRPGREEV-----RVD 132  
Qy 122 DALGRILAE 130  
Db 133 DFLAELTAE 141

RESULT 5  
E84182  
type II secretion system protein [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: E84182  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: E84182  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-548 <STO>  
A:Cross-references: GB:AE004437; NID:g10579863; PIDN:AAG18825.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: gspEI

Query Match 12.0%; Score 84.5; DB 2; Length 548;  
Best Local Similarity 26.68; Pred. No. 1.7;  
Matches 41; Conservative 11; Mismatches 45; Indels 57; Gaps 7;

Qy 11 FRIPVELRYEGD-----PYAIRMTFLPGDAPVTWAFGRELLDGLNSPSPGD-GDVH 62  
Db 184 FDVPVIRYTPGDDADAFITTYAVREPDLPGDDARIAACKDRITWNGVATGSDSDGDH 243  
Qy 63 IGPTPEGLGVHRLQVGADRA---LFRAGTAPLV-----AFLDRTDKLVP 106  
Db 244 AAAT-----ADRARSLARRHTAPKTRHWCGAIADRVREAVADRGPAVPP 288  
Qy 107 LGQEH-----LGDF-----DGNLE 122  
Db 289 VGDDHVADLAYVVRDLVGDAELTIPIRDPNLE 322

RESULT 6  
T19428  
hypothetical protein C24H11.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T19428  
R:Lloyd, C.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19123  
A:Accession: T19428  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A:Residues: 1-261 <WIL>  
A:Cross-references: EMBL:Z81475; PTDN:CA803913.1; GSPDB:GN00021; CESP:C24H11.6  
A:Experimental source: clone C24H11  
C:Genetics:  
A:Gene: CESP:C24H11.6  
A:Map position: 3  
A:Introns: 25/1; 56/3; 126/1; 194/3

Query Match	11.9%	Score 84;	DB 2;	Length 261;
Best Local Similarity	28.7%;	Pred. No. 0.76;		
Matches 33;	Conservative 15;	Mismatches 47;	Indels 20;	Gaps 6;

Qy	22	GDPYAIRMTFHL--PGDAPVTWAFGRLLLDGLNSPSGDGVHIGPTPEEGLGDVHIRLQ	79
		:      - :                             :	
Db	58	GDP----VTSHLLPSGRVPICGHVF-----LRGNGNPVSTDSRHGPV-PEALVOIRLSLR	107

**Qy** 80 VGADRALFRAGTAPLVAFLDRTDKLVPLGQEHT----IGDFDGNLEDALGRILA 129  
:  
||| | : | || : | : | : | : |  
**Db** 108 IWPPE---RAGWTSDHWFVNDKTKLLKMADRMTQLODMINEMAGLMNTAIGVLQA 159

## RESULT 7

A22612  
carbonate dehydratase (EC 4.2.1.1) III - horse  
N/Alternate names: carbonate dehydratase III  
C/Species: Equus caballus (domestic horse)  
C/date: 22-Jul-1987 #sequence\_revision 22-Jul-1987 #text\_change 25-Apr-1997  
C/accession: A22612  
R/Wendorff, K.M.; Nishita, T.; Jabusch, J.R.; Deutsch, H.F.  
J. Biol. Chem. 260, 6129-6132, 1985  
A/title: The sequence of equine muscle carbonic anhydrase.  
A/reference number: A22612; MUID:85207593

A:Accession: A22612  
A:Molecule type: protein  
A:Residues: 1-259 <WEN>  
C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc  
F:4-258/Domain: carbonic anhydrase homology <CAH>  
F:1/Modified site: acetylated amino end (Ala) #status predicted

Query Match	11.9%	Score 83.5;	DB 2;	Length 259;
Best Local Similarity	27.1%	Pred. No. 0.84;		
Matches	29;	Conservative	16;	Mismatches 35;
				Indels 27;
				Gaps 5;

**QY** 24 PYAIRMTFHLPGDAPVTWA----FGRELLDLGLNSPSPGDGVHGP:-----TEPEG 70  
|| :| ||| | | | :|| : || | :| :|  
**Db** 86 PYRURO-FHL-----HWGSSDDHGSEHTVDGVKYAAELHLVWNPKNKNTYGALKOPDG 139

QY 71 LGDVHRLQVGADRALFRAGTAPLVAFLDRTDKLVPLGQEH TLGDFD 117

```
Db      : | : |:| :| : || :|| :|| :||
139 IAVGVFLKIGREKEGFQ-----LF'DALDKIKTKGKEAPFTNFD 178
```

## RESULTS

A69648  
 2-keto-3-deoxygluconate kinase kdgK - *Bacillus subtilis*  
 C:Species: *Bacillus subtilis*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: A69648  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero  
 C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chik  
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallego, J.; Harwood, C.R.H.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Iech, K.T.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinis, Koester, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinis, A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadate, Y.; Sato, T.; Scanlon, Akeuchi, S.; Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.J.; Sekowska, A.; Serod Tanaka, T.; Terpestra, P.; Tononi, A.; Tosato, V.; Uchiyama, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tononi, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A;Reference number: A69580; MUID:98044033

A:Accession: A69648  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-324 <KUN>  
A:Cross-references: GB:299115; GB:AL009126; NID:g2634478; PIDN:CAB14128.1; PIR:1000000000  
C:Experimental source: strain 168  
C:Genetics:  
A:Gene: kdgk  
C:Superfamily: ribokinase

Query Match	11.5%	Score 81;	DB 2;	Length 324;
Best Local Similarity	22.8%;	pred. No. 2;		
Matches	34:	Conservative	26:	Mismatches
			43:	Indels
			46:	Gaps

**QY** 13 IPVELRYEVDPYAIRMTFHL-----PGDAPVTWAFGRELLLDGLNSPSG- 57  
|| || | :| :| : | | : :: :| :

**Dh** 137 TPDAI SAEKND-----FTYHYVNMNRWACKTISPPNVRPSPWP-DOATWVHTINDIAGL 190

[illegible]

09 100 RTDKI.VPI.GOEHTLGD.FDGNLEDAIGRIL 128

249 RYDRW-----DTVGAGDGFVAVGVTSGL 272

## RESULTS

probable transcription regulator MAD-4 - human  
 S71811  
 N:Alternate names: tumor suppressor DPC4  
 C:Species: Homo sapiens (man)  
 C:Date: 12-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 10-Dec-1999  
 C:Accession: S71811  
 R:Zhang, Y.; Feng, X.H.; Wu, R.Y.; Derynck, R.  
 Nature 383, 168-172, 1996  
 A:Title: Receptor-associated Mad homologues synergize as effectors of the TGF-beta re  
 A:Reference number: S71797; MUID:96371046  
 A:Accession: S71811  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-552 <ZHA>  
 C:Superfamily: human transcription regulator MAD-4  
 C:Keywords: transcription regulation

Query Match	11.5%	Score 81;	DB 2;	Length 552;
Best Local Similarity	25.7%	Pred. NO. 3.8;		
Matches	38;	Conservative	16;	Mismatches 28;
				Indels 66;
				Gaps 9

QY 50 DGLNSPSGGD-----VHIG---PTEPEGLGDVHIRLQVGADRAL 86

DH 351 DGVVDSG-GRPEICGQISNVHDTFAFERAPLHTCKGVQICEKGGGVWVRG|--SHAV 407

QY 87 F-----RAGTAPLVAFLDRTDKLP 106  
| | | | |  
408 EVQGVVDFRCPAD-----CDATVETV 421

C  
 T  
 E  
 E  
 C  
 E  
 C

RESULT 10  
AD3237 conserved hypothetical protein Atu6078 [imported] - Agrobacterium tumefaciens (strain C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

A: Molecule type: DNA  
A: Residues: 1-408 <KUR>  
A: Cross-references: GB:AE008690; PID: AAL46314.1; PID: g17744099; GSPDB: GN00189  
A: Experimental source: strain C58 (Dupont)  
C: Genomics:  
A: Gene: Atu6078  
A: Genome: plasmid

```

Query Match      11.28; Score 78.5; DB 2; Length 260;
Best Local Similarity 27.18; Pred. No. 2.6;
Matches 25; Conservative 14; Mismatches 37; Indels 27; Gaps 5;

Qy 24 PYAIRMTFHLPGDAPVWA-----FGRELLLDGLNSPGDGDVHGIC-----TPEG 70
   || || || || || || || || || || || || || || || || || || || ||
Db 87 PYRLRO-FHL-----HWGSSDDHGSHSDVGKYAAAEHLVHWNPKYNTSEEAALKPDG 139

```

```

RESULT 12
S15271
endooglucanase C (EC 3.2.1.-) - Cellulomonas flm1
QY      :  LGDVHHLKQVGADRALRAGAPAPLVAELDRTRDKVPLGQERTLGGDF 117
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 140  IAVVGIFLKTIGREGEFO-----ILLDALDKITKRGKEAPFNHF 179

```

C/Date: 21-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 01-Dec-2000  
C/Accession: S15271; A43636  
R/Coutinho, J.B.; Moser, B.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr., R.C.  
Mol. Microbiol. 5, 1221-1233, 1991  
A/Title: Nucleotide sequence of the endoglucanase C gene (cenC) of *Cellulomonas fimi*,  
A/Reference number: S15271; MUID: 92065819  
A/Accession: S15271

A:Molecule type: DNA  
A:Residues: 1-1101 <COU>  
A:Cross-references: EMBL:X5758; NID:Q48847; PID:G580980  
A:Note: the authors translated the codon GAC for residues 361, 380, 400, 550, 670, and 670, and  
R:Moser, B.; Gilles, N.R.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr., R.C.  
Appl. Environ. Microbiol. 55, 2480-2487, 1989  
A:Title: Purification and characterization of endoglucanase C of Cellulomonas fimi, C  
A:Reference number: A43636; MUID:90103465  
A:Accession: A43636  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-64, 'W' <MOS>

```

C:Cross-references: GB:M29707; NID:G144417; PID:AAA35087.1; PID:G144420
C:Genetics:
A:Gene: cenc
A:Start codon: GTG
C:Keywords: glycosidase; hydrolase

      Query Match      11.2%; Score 78.5; DB 2; Length 1101;
      Best Local Similarity 27.4%; Pred. No. 15;
      Matches 33; Conservative 15; Mismatches 45; Indels 25; Gaps 5;

Qy 28 RMTHLPGDAPVTHAFGRELLLDGLNSPGSDGDVHICGTEPEGLGDVHIRLOV-----GAD 83
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 211 RMCVDLPGGGONGPDAG--LVYNGV--PVGEGESIVLSFTASATPDMPVRVLVGGGGAY 266
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 84 RALFRAGTAPLVAFLDRTDLKVLPGQHTL-----GDFDGNLEDAIGRLAAE 131
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 267 RTAEQGSAPLTG-----EPATREYAFTSNLTFFPDGDGACQVAFHLGKAGAY 315
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 13  
E91295  
peptide chain release factor RF-3 [imported] - *Escherichia coli* (strain O157:H7, subs  
C; Species: *Escherichia coli*  
C; Date: 18-Jul-2001 #sequence revision 18-Jul-2001 #text change 18-Jul-2001

C:Accession: E91295  
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: E91295  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-529 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA038756.1; PID:g13364811; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECS333

Query Match 11.1%; Score 78; DB 2; Length 529;  
Best Local Similarity 26.8%; Pred. No. 7.1;  
Matches 41; Conservative 18; Mismatches 46; Indels 48; Gaps 8;  
QY 1 MSFLVSEELSPRIPVELRYEYVGPYAIRMTFHLPGDAPVTWAFG-----GCAPITWPGCGKLFKGVVHYLYKDET 190  
Db 138 LTFMKNLDRDIPMELLDVENELKI-----GCAPITWPGCGKLFKGVVHYLYKDET 190  
QY 45 -----REL-LLDGLNSPSGDGVHIGTPEPEGLGDVHRLQVGA-----DRALFR 88  
Db 191 YLYQSGKGHTTQEVRIYKGLNPP--DLDAVGEDLAQQLRD-ELELYKGASNEFDKELFL 247  
QY 89 AGTAPLVAFLDRTDKLVPGLQEHITGDFDGNLE 121  
Db 248 AGEITPVFF---GTALGNFGVDHML---DGLVE 274

RESULT 14  
159305  
translation releasing factor RF-3 [validated] - *Escherichia coli*  
N:Alternate names: peptide chain release factor 3; translation termination factor RF-3  
C:Species: *Escherichia coli*  
C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 02-Feb-2001  
R:Accession: 159305; S56599; S63518; F65252; 159306; S39336  
R:Mikuni, O.; Ito, K.; Mofat, J.; Matsumura, K.; McCaughan, K.; Nobukuni, T.; Tate, W.;  
Proc. Natl. Acad. Sci. U.S.A. 91, 5798-5802, 1994  
A:Title: Identification of the prfC gene, which encodes peptide-chain-release factor 3  
A:Reference number: 159305; MUID:94286525  
A:Accession: 159305  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-529 <RES>  
A:Cross-references: GB:D17724; NID:g409376; PIDN:BAA04578.1; PID:g409377  
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
Nucleic Acids Res. 23, 2105-2119, 1995  
A:Title: Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.  
A:Reference number: S6314; MUID:95334362  
A:Accession: S6599  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-529 <BUR>  
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AA97271.1; PID:g537215  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
R:Mortensen, K.K.; Hansen, H.F.; Grentzmann, G.; Buckingham, R.H.; Sperling-Petersen, H.  
Eur. J. Biochem. 234, 732-736, 1995  
A:Title: Osmo-expression and fast two-step purification of *Escherichia coli* translation  
A:Reference number: S63518; MUID:96163454  
A:Accession: S63518  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 2-19 <MOR>  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: F65252

A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-529 <BLAT>  
A:Cross-references: GB:AE000508; GB:U00096; NID:g2367382; PIDN:AA077328.1; PID:g17908  
A:Experimental source: strain K-12, substrain MG1655  
R:Grentzmann, G.; Brechemier-Baey, D.; Heurque, V.; Mora, L.; Buckingham, R.H.  
Proc. Natl. Acad. Sci. U.S.A. 91, 5848-5852, 1994  
A:Title: Localization and characterization of the gene encoding release factor RF3 in  
A:Reference number: 159306; MUID:94286535  
A:Accession: 159306  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-314, 'L', 316-529 <RE2>  
A:Cross-references: EMBL:Z26313; NID:g435273; PIDN:CAA81223.1; PID:g435274  
C:Genetics:  
A:Gene: prfC; tos  
C:Superfamily: translation elongation factor Tu homology  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:14-145/Domain: translation elongation factor Tu homology <ETU>  
F:20-27/Region: nucleotide-binding motif A (P-loop)  
F:142-145/Region: GTP-binding NKXD motif

Query Match 11.1%; Score 78; DB 2; Length 529;  
Best Local Similarity 26.8%; Pred. No. 7.1;  
Matches 41; Conservative 18; Mismatches 46; Indels 48; Gaps 8;  
QY 1 MSFLVSEELSPRIPVELRYEYVGPYAIRMTFHLPGDAPVTWAFG-----GCAPITWPGCGKLFKGVVHYLYKDET 190  
Db 138 LTFMKNLDRDIPMELLDVENELKI-----GCAPITWPGCGKLFKGVVHYLYKDET 190  
QY 45 -----REL-LLDGLNSPSGDGVHIGTPEPEGLGDVHRLQVGA-----DRALFR 88  
Db 191 YLYQSGKGHTTQEVRIYKGLNPP--DLDAVGEDLAQQLRD-ELELYKGASNEFDKELFL 247

QY 89 AGTAPLVAFLDRTDKLVPGLQEHITGDFDGNLE 121  
Db 248 AGEITPVFF---GTALGNFGVDHML---DGLVE 274  
RESULT 15  
G86136  
peptide chain release factor RF-3 [imported] - *Escherichia coli* (strain O157:H7, subs  
C:Species: *Escherichia coli*  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: G86136  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G86136  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-529 <STO>  
A:Cross-references: GB:AE005174; NID:g12519398; PIDN:AAG59555.1; GSPDB:GN00145; UMG:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: prfC

Query Match 11.1%; Score 78; DB 2; Length 529;  
Best Local Similarity 26.8%; Pred. No. 7.1;  
Matches 41; Conservative 18; Mismatches 46; Indels 48; Gaps 8;  
QY 1 MSFLVSEELSPRIPVELRYEYVGPYAIRMTFHLPGDAPVTWAFG-----GCAPITWPGCGKLFKGVVHYLYKDET 190  
Db 138 LTFMKNLDRDIPMELLDVENELKI-----GCAPITWPGCGKLFKGVVHYLYKDET 190  
QY 45 -----REL-LLDGLNSPSGDGVHIGTPEPEGLGDVHRLQVGA-----DRALFR 88  
Db 191 YLYQSGKGHTTQEVRIYKGLNPP--DLDAVGEDLAQQLRD-ELELYKGASNEFDKELFL 247

Search completed: July 18, 2002, 14:17:33  
Job time: 9543 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:16:11 ; Search time 45.63 seconds  
(without alignments)  
114.555 Million cell updates/sec

Title: US-09-749-185-3

Perfect score: 704

Sequence: 1 MSFLVSEELSFRIPELVRYE.....FDGNLEDALGRILAEQNAQ 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	83.5	11.9	259	1 CAH3_HORSE	P07450 equus cabal
2	82	11.6	551	1 SMA4_MOUSE	P97471 mus musculu
3	81	11.5	324	1 KDGG_BAGSU	P50845 bacillus su
4	81	11.5	552	1 SMA4_HUMAN	Q13485 homo sapien
5	81	11.5	552	1 SMA4_PIG	Q9gkq9 sus scrofa
6	81	11.5	552	1 SMA4_RAT	Q70437 rattus norv
7	78.5	11.2	259	1 CAH3_RAT	P14141 rattus norv
8	78.5	11.2	1101	1 GUNC_CELFI	P14090 cellulomona
9	78	11.1	528	1 RF3_ECOLI	P33998 escherichia
10	78	11.1	690	1 VATI_DEIRA	Q9rwh3 deinococcus
11	77.5	11.0	259	1 CAH3_MOUSE	P16015 mus musculu
12	77	10.9	795	1 LON2_AQUAE	O65603 aquifex ao
13	77	10.9	826	1 LON2_MYXAA	P36774 myxococcus
14	75.5	10.7	1046	1 CHID_VIBFU	P96156 vibrio furn
15	74	10.5	941	1 LON2_ARATH	P93655 arabidopsis
16	73.5	10.4	432	1 YQ31_MYCTU	P71930 mycobacteri
17	73.5	10.4	484	1 HXK1_SCHPO	Q09756 schizosacch
18	73	10.4	528	1 RF3_SALTY	Q56121 salmonella
19	72.5	10.3	338	1 C1F_PHOLA	P95522 phormidium
20	72.5	10.3	339	1 ENDA_HALVO	O07118 halobacteri
21	72	10.2	323	1 GSHB_SYNP7	Q32463 synechococc
22	72	10.2	980	1 PEX6_HUMAN	Q13608 homo sapien
23	72	10.2	3591	1 PHAB_BORPE	P12255 bordetella
24	71.5	10.2	706	1 GYS_NEUCR	O93869 neurospora
25	71	10.1	837	1 STA6_MOUSE	P52633 mus musculu
26	71	10.1	964	1 LON2_MAIZE	P93648 zea mays (m
27	71	10.1	978	1 PEX6_RAT	P54777 rattus norv
28	69.5	9.9	259	1 CAH3_HUMAN	P07451 homo sapien
29	69.5	9.9	364	1 DAP1_ARATH	Q91zx6 arabidopsis
30	69.5	9.9	373	1 C1S2_MYCTU	Q10529 mycobacteri
31	68	9.7	401	1 SY1_HAEIN	P43836 haemophilus
32	68	9.7	779	1 LON2_BACBR	P36772 bacillus br
33	67.5	9.6	650	1 DMWD_MOUSE	Q08274 mus musculu

34	67.5	9.6	1370	1 Z261_HUMAN	Q14202 homo sapien
35	67	9.5	143	1 YW84_MYCTU	P96889 mycobacteri
36	67	9.5	396	1 NUD1_RHIME	P56907 rhizobium m
37	67	9.5	527	1 RF3_HAEIN	P43928 haemophilus
38	67	9.5	597	1 CCBA_RAT	P54283 rattus norv
39	66.5	9.4	292	1 AQP3_RAT	P47862 rattus norv
40	66.5	9.4	411	1 CSD_AERPE	Q9yab6 aeropyrum p
41	66.5	9.4	572	1 TCMG_STRGA	P39888 streptomyc
42	66	9.4	201	1 YJ10_MYCTU	O07723 mycobacteri
43	66	9.4	220	1 YP11_ECTHA	P42515 ectothiorho
44	66	9.4	343	1 HRCAL_MYCTU	O05824 mycobacteri
45	66	9.4	423	1 SYH_MYCTU	Q50641 mycobacteri

ALIGNMENTS

RESULT 1  
CAH3\_HORSE  
ID CAH3\_HORSE STANDARD; PRT; 259 AA.  
AC P07450:  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Carbonic anhydrase III (EC 4.2.1.1) (Carbonate dehydratase III) (CA-  
DE III).  
GN CA3.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=85207593; PubMed=3922970;  
RA Wendorff K.M., Nishita T., Jabusch J.R., Deutsch H.F.;  
RT "The sequence of equine muscle carbonic anhydrase.";  
J. Biol. Chem. 260:6129-6132(1985).  
CC -!- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.  
CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE  
CC FAMILY.  
CC PIR: A22612; A22612.  
DR HSSP; P00918; 1BV3.  
DR InterPro: IPR001148; Carb.anhydrase.  
DR Pfam: PF00194; carb.anhydrase; 1.  
DR ProDom: PD000865; Carb.anhydrase; 1.  
DR PROSITE; PS00162; EUK\_CO2\_ANHYDRASE; 1.  
KW Lyase; Zinc; Acetylation.  
FT MOD\_RES 1 ACETYLATION.  
FT METAL 93 ZINC (CATALYTIC).  
FT METAL 95 ZINC (CATALYTIC).  
FT METAL 118 ZINC (CATALYTIC).  
SQ SEQUENCE 259 AA; 29380 MW; B3EECDF89BFED02 CRC64;

Query Match 11.9%; Score 83.5; DB 1; Length 259;  
Best Local Similarity 27.1%; Pred. No. 0.27; 35; Indels 27; Gaps 5;  
Matches 29; Conservative 16; Mismatches 27;

Qy	24	PYAIRMTFLPGDAPVTWA-----FGRELLDGLNPSGDGVHIGP-----TEPEG 70
Db	86	PYRLRQ-PHL-----HWGSSDDHSGEHTVDGVKAAELHLVHWNPKYNTYTGALKQPDG 138
Qy	71	IGDVHIRLQVGADRALFRAGTAPLVAFLDRTDKLVLPGOEHTLGDFD 117
Db	139	IATVGVFLKIGKEGEFQ-----LFLDALDKIKTKGKEAPTTFD 178

RESULT 2  
SMA4\_MOUSE  
ID SMA4\_MOUSE STANDARD; PRT; 551 AA.  
AC P97471; Q9CW56;





CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; AF045447; AAC03051.1; -  
DR EMBL; AF045438; AAC03051.1; JOINED.  
DR EMBL; AF045439; AAC03051.1; JOINED.  
DR EMBL; AF045440; AAC03051.1; JOINED.  
DR EMBL; AF045441; AAC03051.1; JOINED.  
DR EMBL; AF045442; AAC03051.1; JOINED.  
DR EMBL; AF045443; AAC03051.1; JOINED.  
DR EMBL; AF045444; AAC03051.1; JOINED.  
DR EMBL; AF045445; AAC03051.1; JOINED.  
DR EMBL; AF045446; AAC03051.1; JOINED.  
DR EMBL; U44378; AAA91041.1; -  
DR PDB; 1YGS; 08-NOV-98.  
DR PDB; 1DD1; 24-NOV-99.  
DR PDB; 1G88; 29-NOV-00.

DR MIM; 600993; -  
DR InterPro; IPR001132; DwarfIn.  
DR InterPro; IPR003619; DwarfIn\_A.  
DR Pfam; PF00968; DwarfIn; 1.  
DR SMART; SM00523; DWA; 1.  
DR SMART; SM00524; DWB; 1.  
KW Transcription regulation; Phosphorylation; Multigene family;  
KW Disease mutation; 3D-structure.  
FT DOMAIN 31 140 DWA.  
FT DOMAIN 275 320 SAD.  
FT DOMAIN 321 530 DWB.  
FT DOMAIN 451 466 POLY-ALA.  
FT VARIANT 493 493 D -> H (IN PANCREATIC CARCINOMA).  
FT /FTID=VAR\_011380.  
SQ SEQUENCE 552 AA; 60439 MW; 7EE3C4647712DA90 CRC64;

Query Match 11.5%; Score 81; DB 1; Length 552;  
Best Local Similarity 25.7%; Pred. No. 1.2;  
Matches 38; Conservative 16; Mismatches 28; Indels 66; Gaps 9;

QY 5 VSELSFRIPV-----ELRYEVDGPYAIRMTFHLPGDAPVYTWAFGREL 49  
| | | | | : | | | | : | | | | : | | | | :  
Db 304 VHNELAFQPPISNHPAPEYWCSTAYFEMDVQVGE-----TFKVPSSCPI-----VTV 350  
QY 50 DGLNSPSGDG-----VHTG-----PTEPEGLGDVHRLQVGADRAL 86  
| | | | | : | | | | : | | | | : | | | | :  
Db 351 DGYVDPSG-GDRFCLGQLSNVHRTEAERLHIGKGVQLECKEGDVWVRCL--SDHAV 407  
QY 87 F-----RAGTAPLVAFLDRTDKLVP 106  
| | | | | : | | | | : | | | | : | | | | :  
Db 408 FVQSYLDREAGR-----GDVHKIYP 431

RESULT 5  
SMA4\_PIG STANDARD; PRT; 552 AA.  
ID SMA4\_PIG  
AC Q9GK09  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mothers against decapentaplegic homolog 4 (SMAD 4) (Mothers against  
DE DPP homolog 4) (Smad4).  
GN MADH4 OR SMAD4.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Tissue=Muscle;  
RA Ito Y., Awata T.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
CC -I- FUNCTION: COMMON MEDIATOR OF SIGNAL TRANSDUCTION BY TGF-BETA  
(TRANSFORMING GROWTH FACTOR) SUPERFAMILY; SMAD4 IS THE COMMON

CC SMAD (CO-SMAD) (BY SIMILARITY).  
CC -I- SUBUNIT: MAY FORM TRIMERS WITH RECEPTOR-REGULATED SMAD (R-SMAD).  
CC INTERACTS WITH C-SKI, MSG1 AND ATF2 (BY SIMILARITY).  
CC -I- SUBCELLULAR LOCATION: IN THE CYTOPLASM IN THE ABSENCE OF LIGAND;  
CC MIGRATION TO THE NUCLEUS WHEN COMPLEXED WITH R-SMAD (BY  
CC SIMILARITY).  
CC -I- SIMILARITY: BELONGS TO THE DWARFIN/SMAD FAMILY.  
CC -I- SIMILARITY: CONTAINS 1 DWA/MH1 DOMAIN.  
CC -I- SIMILARITY: CONTAINS 1 DWB/MH2 DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB053483; BAB20909.1; -  
DR InterPro; IPR001132; DwarfIn.  
DR InterPro; IPR003619; DwarfIn\_A.  
DR Pfam; PF00968; DwarfIn; 1.  
DR SMART; SM00523; DWA; 1.  
DR SMART; SM00524; DWB; 1.  
KW Transcription regulation; Multigene family.  
KW Disease mutation; 3D-structure.  
FT DOMAIN 31 140 DWA.  
FT DOMAIN 275 320 SAD.  
FT DOMAIN 321 530 DWB.  
FT DOMAIN 451 466 POLY-ALA.  
SQ SEQUENCE 552 AA; 60415 MW; 4304796DE7571CB3 CRC64;

Query Match 11.5%; Score 81; DB 1; Length 552;  
Best Local Similarity 25.7%; Pred. No. 1.2;  
Matches 38; Conservative 16; Mismatches 28; Indels 66; Gaps 9;

QY 5 VSELSFRIPV-----ELRYEVDGPYAIRMTFHLPGDAPVYTWAFGREL 49  
| | | | | : | | | | : | | | | : | | | | :  
Db 304 VHNELAFQPPISNHPAPEYWCSTAYFEMDVQVGE-----TFKVPSSCPI-----VTV 350  
QY 50 DGLNSPSGDG-----VHTG-----PTEPEGLGDVHRLQVGADRAL 86  
| | | | | : | | | | : | | | | : | | | | :  
Db 351 DGYVDPSG-GDRFCLGQLSNVHRTEAERLHIGKGVQLECKEGDVWVRCL--SDHAV 407  
QY 87 F-----RAGTAPLVAFLDRTDKLVP 106  
| | | | | : | | | | : | | | | : | | | | :  
Db 408 FVQSYLDREAGR-----GDVHKIYP 431

RESULT 6  
SMA4\_RAT STANDARD; PRT; 552 AA.  
ID SMA4\_RAT  
AC Q70437  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mothers against decapentaplegic homolog 4 (SMAD 4) (Mothers against  
DE DPP homolog 4) (Smad4).  
GN MADH4 OR SMAD4.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Tissue=Skeletal muscle;  
RA Miyakita A., Okuno S., Watanabe T.K., Oga K., Tsuji A., Hishigaki H.,  
RA Suto T., Nakagawa K., Nakahara Y., Higashi K.;  
RL "Molecular cloning of rat Smad4 gene."  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20134216; PubMed=10670756;

```
RA Osaki M., Tsukazaki T., Ono N., Yonekura A., Hirota Y., Miyazaki Y.,
RA "China cloning and chromosomal mapping of rat Smad2 and Smad4 and their
RT expression in cultured rat articular chondrocytes.";
RL Endocr. J. 46:695-701(1999).
CC -!- FUNCTION: COMMON MEDIATOR OF SIGNAL TRANSDUCTION BY TGF-BETA
CC (TRANSFORMING GROWTH FACTOR) SUPERFAMILY; SMAD4 IS THE COMMON
CC SMAD (CO-SMAD) (BY SIMILARITY).
CC -!- SUBUNIT: MAY FORM PRIMERS WITH RECEPTOR-REGULATED SMAD (R-SMAD).
CC INTERACTS WITH C-SKI, MSK1 AND ATF2 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: IN THE CYTOPLASM IN THE ABSENCE OF LIGAND;
CC MIGRATION TO THE NUCLEUS WHEN COMPLEXED WITH R-SMAD (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DWARFIN/SMAD FAMILY.
CC -!- SIMILARITY: CONTAINS 1 DWA/MH1 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DWA/MH2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB010954; BAA83092.1; -.
DR EMBL; AF056002; AAC12781.1; -.
DR HSSP: O13485; 1DD1.
DR InterPro: IPR001132; DwarfIn.
DR Pfam: IPR003619; DwarfIn_A.
DR SMART; SM00968; DwarfIn; 1.
DR SMART; SM00523; DWA; 1.
DR SMART; SM00524; DWB; 1.
KW Transcription regulation; Multigene family.
FT DOMAIN 31 140 DWA.
FT DOMAIN 275 320 SAD.
FT DOMAIN 321 530 DWB.
FT DOMAIN 451 466 POLY-ALA.
SQ SEQUENCE 552 AA: 60469 MW; 7AE0540AB4DF0E77 CRC64;

Query Match 11.5%; Score 81; DB 1; Length 552;
Best Local Similarity 25.7%; Pred. NO. 1.2;
Matches 38; Conservative 16; Mismatches 28; Indels 66; Gaps 9;

QY 5 VSEELSFIPV-----ELRYEYVGPYAIRMTFLPGDAPVTWAFGRELLL 49
DB 304 VHNELAFQPPISNHPAPEYWCISAIYFEMDVQGE-----TPKVPSSCPI-----VTV 350
QY 50 DGLNSPSGDG-----VHIG-----PTEPEGLGDVHIRLQVGGADRAL 86
DB 351 DGVVDPSG-GDRFCLGQLSNVHRTAIEARLHIGKGVQLECKGEGVWVRCI--SDHAV 407
QY 87 F-----RAGTAPLVAFLDRTKLVP 106
DB 408 FVQSYLDREAGR-----GDVHVIYP 431

RESULT 7
CAH3_RAT STANDARD; PRT; 259 AA.
AC P14141; O54961;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carbonic anhydrase III (EC 4.2.1.1) (carbonate dehydratase III) (CA-
DE III).
GN CA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE-8916882; PubMed-2852973;
RA Kelly C.D., Carter N.D., Jeffery S., Edwards Y.H.;
RT "Characterisation of cDNA clones for rat muscle carbonic anhydrase
RT III.";
RL Biosci. Rep. 8:401-406(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Mallis R.J., Jung C.-H., Chatterjee T.K., Fisher R.A., Thomas J.A.;
RT "cDNA sequence of rat liver carbonic anhydrase III.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Liver;
RX MEDLINE-20361776; PubMed-10900145;
RA Ikeda M., Ishii Y., Kato H., Akazawa D., Hatsumura M., Ishida T.,
RA Matsusue K., Yamada H., Oguri K.;
RT "Suppression of carbonic anhydrase III in rat liver by a
RT dioxin-related toxic compound, coplanar polychlorinated biphenyl,
RT 3,3',4,4',5-pentachlorobiphenyl.";
RL Arch. Biochem. Biophys. 380:159-164(2000).
CC -!- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.
CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22413; AAA40846.1; -.
DR EMBL; AF037072; AAB92558.1; -.
DR EMBL; AB030829; BAB08111.1; -.
DR HSSP: P00918; 1CIM.
DR InterPro: IPR001148; Carb_anhydrase.
DR Pfam: PF00194; carb_anhydrase; 1.
DR ProDom; PD000865; Carb_anhydrase; 1.
DR PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1.
KW Lyase; Zinc.
FT INIT_MET 0 0
FT METAL 93 93 ZINC (CATALYTIC).
FT METAL 95 95 ZINC (CATALYTIC).
FT METAL 118 118 ZINC (CATALYTIC).
FT CONFLICT 7 7 A -> G (IN REF. 1).
FT CONFLICT 129 130 FG -> SE (IN REF. 1).
FT CONFLICT 223 224 KL -> NV (IN REF. 1).
SQ SEQUENCE 259 AA: 29300 MW; 7BF7DA65E9493F21 CRC64;

Query Match 11.2%; Score 78.5; DB 1; Length 259;
Best Local Similarity 27.1%; Pred. No. 0.87;
Matches 29; Conservative 14; Mismatches 14; Indels 27; Gaps 5;

QY 24 PYAIRMTFLPGDAPVTWA-----FGRELLDGLNSPSGDGVHIGP-----TEPEG 70
DB 86. PYRLRQ-PHL-----HWGSSDDHGHSEHTVGVKVAALHVLHWNPKYNTFGEALKQPDG 138
QY 71 LQGVHRLQVGGADRALFRAGTAPLVAFLDRTKLVLPGQEHITLGDFF 117
DB 139. IAVVGIFLKGKGEFQ-----ILLDALDKIKTKGKAPFNHFD 178

RESULT 8
GUNC_CELFI
ID GUNC_CELFI STANDARD; PRT; 1101 AA.
AC P14090;
DT 01-JAN-1990 (Rel. 13, Created)
```



RT factor RF3 in Escherichia coli.;"

RL Proc. Natl. Acad. Sci. U.S.A. 91:5848-5852(1994).

RN [3]

RC SEQUENCE FROM N.A.

RX STRAIN-K12 / MG1655;

RA MEDLINE-95334362; PubMed=7610040;

RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,

RA Blattner F.R.;

RT "Analysis of the Escherichia coli genome VI: DNA sequence of the

RT region from 92.8 through 100 minutes.;"

RL Nucleic Acids Res. 23:2105-2119(1995).

RN [4]

RC SEQUENCE FROM N.A.

RX STRAIN-O157:H7 / EDL933 / ATCC 700927;

RC MEDLINE-21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocous K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"

RL Nature 409:529-533(2001).

RN [5]

RC SEQUENCE FROM N.A.

RX STRAIN-O157:H7 / RIMD 0509952;

RX MEDLINE-21156231; PubMed=11256796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,

RA Kubara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli

RT O157:H7 and genomic comparison with a laboratory strain K-12.;"

RL DNA Res. 8:11-22(2001).

RN [6]

RC SEQUENCE OF 1-18.

RX MEDLINE-96163454; PubMed=8575429;

RA Mortensen K.K., Hansen H.F., Grentzmann G., Buckingham R.H.,

RA Sperling-Petersen H.U.;

RT "Osmo-expression and fast two-step purification of Escherichia coli

RT translation termination factor RF-3.;"

RL Eur. J. Biochem. 234:732-736(1995).

RN [7]

RC CHARACTERIZATION.

RX MEDLINE-95256223; PubMed=7737996;

RA Grentzmann G., Brechemier-Baey D., Heurque-Hamard V., Buckingham R.H.;

RT "Function of polypeptide chain release factor RF-3 in Escherichia

RT coli. RF-3 action in termination is predominantly at UGA-containing

RT stop signals.;"

RL J. Biol. Chem. 270:10595-10600(1995).

CC -!- FUNCTION: INCREASES THE FORMATION OF RIBOSOMAL TERMINATION

CC COMPLEXES AND STIMULATES ACTIVITIES OF RF-1 AND RF-2. IT BINDS

CC GUANINE NUCLEOTIDES AND HAS STRONG PREFERENCE FOR UGA STOP CODONS.

CC IT MAY INTERACT DIRECTLY WITH THE RIBOSOME. THE STIMULATION OF RF-

CC 1 AND RF-2 IS SIGNIFICANTLY REDUCED BY GTP AND GDP, BUT NOT BY

CC GMP.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.

CC PRFC SUBFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; D17724; BAA04578.1; -

DR EMBL; 226313; CA881223.1; -

DR EMBL; U14003; AAA97271.1; -

DR EMBL; AE000508; AAC77328.1; -

DR EMBL; AE005668; AAG59555.1; -

DR

DR EMBL; AP002569; BAB38756.1; -

DR EcoGene; EG12114; prfC.

DR InterPro; IPR000795; GTP\_EFTU.

DR InterPro; IPR004161; GTP\_EFTU\_D2.

DR Pfam; PF00009; GTP\_EFTU; 1.

DR Pfam; PF03144; GTP\_EFTU\_D2; 1.

DR PROSITE; PS00301; EFATOR\_GTP; 1.

KW Protein biosynthesis; GTP-binding; Complete proteome.

FT INIT\_MET 0 0

FT NP\_BIND 19 26 GTP (BY SIMILARITY).

FT NP\_BIND 87 91 GTP (BY SIMILARITY).

FT NP\_BIND 141 144 GTP (BY SIMILARITY).

FT CONFLICT 314 314 V -> L (IN REF. 2).

SQ SEQUENCE 528 AA; 59443 MW; 3EE94DE2B4C6B1C6 CRC64;

Query Match 11.1%; Score 78; DB 1; Length 528;

Best Local Similarity 26.8%; Pred.No. 2.3;

Matches 41; Conservative 18; Mismatches 46; Indels 48; Gaps 8;

QY 1 MSFLVSEELSPRIPELVRYEGDYPVIAIRTFHLPDGPVTFWAFG-----GCAPITWFIGCKLFKGVHYLYKDTF 189

DB 137 LTFMKNLDRDIRDPWELLDEVENELKI-----CCAPITWFIGCKLFKGVHYLYKDTF 189

QY 45 -----REL-LLDGLNSPSGSDGVHIGTEPEGLDVHRIQVGA-----DRALFR 88

DB 190 YLYQSGKGHTIQEVRIVKGLNPN--DLDAVGEDLAQQLRD-ELELVKSGASNEFDKELF 246

QY 89 AGTAELVAFLDRTDKLVPLGQEHTELTGDFGNLE 121

DB 247 AGEITPVFF--GTALGNFGVDHML---DGLVE 273

RESULT 10

VATL\_DEIRA STANDARD; PRT; 690 AA.

AC Q9RWH3;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE V-type ATP synthase subunit I (EC 3.6.3.14) (V-type ATPase subunit I).

GN ATP1 OR DR0695.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

OX NCBI\_TaxID=1299;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-R1;

RX MEDLINE-20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,

RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans R1.;"

RL Science 286:1571-1577(1999).

CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON

CC GRADIENT ACROSS THE MEMBRANE.

CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +

CC H(+) (Out).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----



Query Match	11.0%	Score 77.5	DB 1	Length 259	
Best Local Similarity	27.1%	Pred. No. 1.1			
Matches	29	Conservative 14	Mismatches 37	Indels 27	Gaps 5
QY	24	PYAIRMTFHLPGDAPVTWA	-----FGRLLLDGLNSGSGDGVHGP	-----TPEEG 70	
DB	86	PYRLRQ-FHL	-----HWSSDDHSGSEHTVDGVKYAAELHLVHWNPRYNTFGKALKOPDG 138		
QY	71	LGDVHRLQVGADRALFRACFAPLVAFLDRDTDKLVPLGQEHHTLGDFFD 117			
DB	139	IAVVGILLIKRGKGFQ	-----ILLDALDKIKYKKEAPFTHFD 178		
RESULT 12					
LON_AQAE		STANDARD;	PRT;	795 AA.	
ID	AC	066805;			
DT	15-DEC-1998	(Rel. 37, Created)			
DT	15-DEC-1998	(Rel. 37, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	ATP-dependent	protease La (EC 3.4.21.53).			
GN	LON OR AQ_242				
OS	Aquifex aeolicus.				
OC	Bacteria; Aquificales; Aquificaceae; Aquifex.				
OX	NCBI_TaxID=63363;				
RN	[1]				
RN	SEQUENCE FROM N.A.				
RC	STRAIN=VF5;				
RC	MEDLINE=98196666; PubMed=9537320;				
RA	Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,				
RA	Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,				
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;				
RT	"The complete genome of the hyperthermophilic bacterium Aquifex				
RT	aeolicus".				
RL	Nature 392:353-358(1998).				
CC	-I- FUNCTION: DEGRADATES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS IN				
CC	PRESENCE OF ATP. HYDROLYZES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED				
CC	IN THE PROTEIN SUBSTRATE (BY SIMILARITY).				
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,				
CC	casein and denaturated serum albumin, in presence of ATP.				
CC	-I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).				
CC	-I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).				
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE				
CC	LON FAMILY OF ATP-DEPENDENT PROTEASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
CC	EMBL; AE000680; AAC06568.1; -				
CC	MEROPS; S16.001; -				
DR	DR	InterPro; IPR003593; AAA.			
DR	DR	InterPro; IPR003959; AAA_subfam.			
DR	DR	InterPro; IPR003111; LON.			
DR	DR	InterPro; IPR001984; Lon_endopep.			
DR	DR	Pfam; PF00004; AAA; 1.			
DR	DR	Pfam; PF02190; LON; 1.			
DR	DR	PRINTS; PR00830; ENDOLAPTASE.			
DR	DR	SMART; SM00382; AAA; 1.			
DR	DR	SMART; SM00464; LON; 1.			
DR	DR	PROSITE; PS01046; LON_SER; 1.			
KW	Hydrolase; Serine protease; ATP-binding; Complete proteome.				
FT	NP_BIND	370	ATP (POTENTIAL).		
FT	ACT_SITE	692 692	BY SIMILARITY.		
FT	SEQUENCE.	795 AA; 89971 MW; EC8FDBB7F24EFEFB CRC64;			
QY	Query Match	10.9%	Score 77	DB 1	Length 795
Best Local Similarity	27.7%	Pred. No. 4.7			



[illegible]

DR PROSITE; PS01095; CHITINASE\_18; FALSE\_NEG.  
KW Hydrolase; Glycosidase; Signal; Periplasmic.  
FT SIGNAL 1 30  
FT CHAIN 31 1046 CHITODEXTRINASE.  
SQ SEQUENCE 1046 AA; 112380 MW; 40680F1642D55A1F CRC64;

Query Match 10.7%; Score 75.5; DB 1; Length 1046;  
Best Local Similarity 30.5%; Pred. No. 9.2;  
Matches 25; Conservative 8; Mismatches 30; Indels 19; Gaps 3;

QY 9 LSRFRPVELRYEGDPAIRMTHTPLPGDAP-----VTWAFGREL-----LLDGL 52  
Db 920 VAFSLPAWKTLPGADTVELDMVYLPISGPANYSVNINGVDYAFKFPQDPLDLADLSSGN 979  
QY 53 NSPSGDDGVHIGTPEPEGLGDV 74  
Db 980 GGGTGGDGTGGTTEP---GDV 998

RESULT 15

LON2\_ARATH STANDARD; PRT; 941 AA.  
AC P93635;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Lon protease homolog 2, mitochondrial precursor (EC 3.4.21.-).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Sarria R., Lyznik A., MacKenzie S.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE LON  
FAMILY OF ATP-DEPENDENT PROTEASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U88087; AAB48000.1; -  
DR InterPro: IPR003593; AAA.  
DR InterPro: IPR003959; AAA\_subfam.  
DR InterPro: IPR003111; LON.  
DR InterPro: IPR001984; Lon\_endopep.  
DR Pfam: PF00004; AAA; 1.  
DR Pfam: PF02190; LON; 1.  
DR PRINTS: PR00830; ENDOLAPTASE.  
DR SMART: SM00382; AAA; 1.  
DR SMART: SM00464; LON; 1.  
DR PROSITE; PS01046; LON\_SER; 1.  
KW Hydrolase; Serine protease; ATP-binding; Multigene family;  
KW Mitochondrion; Transit peptide.  
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).  
FT CHAIN ? 941 LON PROTEASE HOMOLOG 2.  
FT NP\_BIND 464 471 ATP (POTENTIAL).  
FT ACT\_SITE 842 842 BY SIMILARITY.  
SQ SEQUENCE 941 AA; 104101 MW; 2B1FAF5184413C3C CRC64;

Query Match 10.5%; Score 74; DB 1; Length 941;  
Best Local Similarity 31.4%; Pred. No. 11;  
Matches 33; Conservative 8; Mismatches 34; Indels 30; Gaps 5;

QY 52' LNSPSGDDGVHIGT-----EPEGLGDV-----HIRLOVGA-----DRALFRA 89  
Db 462' LSGPPPGVGKTSIGRSIARALNRKFFRFSGVLADVAEIQGHRTTYVGAMPKMKVQCLKSV 521  
QY 90' GTAPLVAFLDRTDKLVPLGQEHHTLGDFDGNLEDAIGRILAEQNA 134  
Db 522' GTANPLVLIDEIDK---LGRGHA-----GDPASALLELDPEQNA 558

Search completed: July 18, 2002, 14:31:53  
Job time: 942 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:15:11 ; Search time 140.15 Seconds  
(without alignments)  
166.638 Million cell updates/sec

Title: US-09-749-185-3

Perfect score: 704

Sequence: 1 MSFLVSEELSFRIPELVRYE.....FDGNLEDALGRILAEQNAQ 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	704	100.0	136	2 P95753	P95753 streptomyc
2	606	86.1	145	2 Q9F9B5	Q9F9B5 streptococ
3	556	79.0	135	2 Q9F9B7	Q9F9B7 streptomyc
4	546	77.6	136	2 Q9X9U2	Q9X9U2 streptomyc
5	464	65.9	135	2 Q9F9B6	Q9F9B6 streptomyc
6	249	35.4	159	2 Q9L268	Q9L268 streptomyc
7	219	31.1	142	2 Q9S2F7	Q9S2F7 streptomyc
8	190	27.0	138	2 Q9X7M8	Q9X7M8 streptomyc
9	150	21.3	156	2 Q9FC07	Q9FC07 streptomyc
10	140	19.9	142	2 Q9X7R1	Q9X7R1 streptomyc
11	99.5	14.1	126	2 Q9RKC9	Q9RKC9 streptomyc
12	84.5	12.0	548	17 Q9HSI1	Q9HSI1 halobacteri
13	84	11.9	261	5 Q9XVD2	Q9XVD2 caenorhabd
14	83.5	11.9	456	10 Q9SR03	Q9SR03 arabidopsis
15	81	11.5	402	2 Q93JF9	Q93JF9 streptomyc
16	81	11.5	436	4 Q9BYG6	Q9BYG6 homo sapien

17	81	11.5	549	13	Q9W639	Q9W639 xenopus lae
18	79	11.2	353	13	Q919P4	Q919P4 brachydanio
19	79	11.2	544	13	Q90Y99	Q90Y99 cyprinus ca
20	79	11.2	547	13	Q90YAL	Q90YAL cyprinus ca
21	78	11.1	547	13	Q90YAO	Q90YAO cyprinus ca
22	78	11.1	700	2	Q34003	Q34003 rhodobacter
23	77.5	11.0	913	2	O54182	O54182 streptomyc
24	77	10.9	544	13	Q90Y98	Q90Y98 cyprinus ca
25	77	10.9	819	10	Q9SNA4	Q9SNA4 arabidopsis
26	76.5	10.9	260	11	Q9ERN8	Q9ERN8 mus musculu
27	76	10.8	762	2	Q9RL29	Q9RL29 streptomyc
28	75.5	10.7	902	16	Q91742	Q91742 pseudomonas
29	75	10.7	381	2	Q9KXV7	Q9KXV7 streptomyc
30	75	10.7	644	2	Q9RJK2	Q9RJK2 streptomyc
31	75	10.7	760	16	Q916K5	Q916K5 pseudomonas
32	75	10.7	1980	5	Q9NDY5	Q9NDY5 leishmania
33	74	10.5	924	10	Q9M9L8	Q9M9L8 arabidopsis
34	74	10.5	981	11	Q99LC9	Q99LC9 mus musculu
35	74	10.5	1235	4	O95428	O95428 homo sapien
36	73.5	10.4	346	16	O34788	O34788 bacillus su
37	73.5	10.4	418	11	Q9ER10	Q9ER10 microtus br
38	73.5	10.4	538	16	Q9RU41	Q9RU41 deinococcus
39	73.5	10.4	1191	12	Q9E6P0	Q9E6P0 turkey heip
40	73	10.4	374	16	Q92Y75	Q92Y75 rhizobium m
41	73	10.4	589	10	Q9C6Z1	Q9C6Z1 arabidopsis
42	73	10.4	674	5	Q9VLJ1	Q9VLJ1 drosophila
43	73	10.4	2034	2	Q93NX7	Q93NX7 streptomyc
44	72.5	10.3	512	17	O26636	O26636 methanother
45	72.5	10.3	1096	10	O04954	O04954 arabidopsis

## ALIGNMENTS

RESULT 1

ID P95753 PRELIMINARY; PRT; 136 AA.  
AC P95753;  
DT 01-MAY-1997 (Tremblrel. 03, Created)  
DE 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE SSGA.  
OS Streptomyces griseus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_taxid=1911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B2682;  
RA Shinichi K., Ensign J.C.;  
RT "Cloning and characterization of a gene involved in sporulation and cell division of Streptomyces griseus.";  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B2682;  
RA Kawamoto S., Ensign J.C.;  
RT "Isolation of mutants of Streptomyces griseus that sporulate in nutrient rich media.";  
RL Nippon Hosenkin Gakkaishi 9:124-135(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B2682;  
RA Kawamoto S., Ensign J.C.;  
RT "Cloning and characterization of a gene involved in regulation of sporulation and cell division of Streptomyces griseus.";  
RL Nippon Hosenkin Gakkaishi 9:136-151(1995).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B2682;  
RX MEDLINE=97286526; PubMed=9141673;  
RA Kawamoto S., Watanabe H., Hesketh A., Ensign J.C., Ochi K.;  
RT "Expression analysis of a ssgA gene product associated with





```
Db 77 VRWPSKTEGSRVVVLVALSSPDGALLLEAPTPQVSAMLERTLRAVPPGTEGAQLG----- 131
Qy 120 LEDALGRILA 129
   : : | : | |
Db 132 IDDGIAELLA 141

RESULT 8
Q9X7M8 PRELIMINARY; PRT; 138 AA.
AC Q9X7M8:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 13, Last annotation update)
DE PUTATIVE REGULATOR.
GN SC5F2A.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
[1]
RN SEQUENCE FROM N.A.
RP Oliver K., Harris D.;
RC STRAIN=A3(2);
RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RC STRAIN=A3(2);
RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL049587; CAB40672.1; -.
SQ SEQUENCE 138 AA; 15261 MW; F531BA295144ABA5 CRC64;

Query Match 27.0%; Score 190; DB 2; Length 138;
Best Local Similarity 31.3%; Pred. No. 3.7e-11;
Matches 46; Conservative 19; Mismatches 46; Indels 36; Gaps 4;

Qy 1 MSFLVSELSFR-----IPVELRYEGDPAIRMTFHLPG---DAPVTWAFQREL 47
   | : : : | | | | : : | : | : | | | | |
Db 1 MSTVIEQSVEARLVAAAPRMPSPATLHYDRADPFVVRMTFPAPATLEGVEVCWTFSSREL 60
   | : : : | | | | : : | : | : | | | | |
Qy 48 LLDGLNSPGDGDVHIGP-----TEPEGLGDVHIRLQVGADRALFRAGTAPLVAF 97
   | : : : | | | | : : | : | : | | | | |
Db 61 LIAGMQEPNGHGVRRPYAYDRTVLFEHFAPEGTAVIHR-----SGELRRF 107
   | : : : | | | | : : | : | : | | | | |
Qy 98 LDRTDKLVPLQGEHTLGDGFDGNLEDAL 124
   | : : : | | | | : : | : | : | | | | |
Db 108 LQAAGELVPVGLHQLDLDHDLAELM 134

RESULT 9
Q9FC07 PRELIMINARY; PRT; 156 AA.
AC Q9FC07:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PUTATIVE REGULATOR.
GN SC8A11.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
```

```

RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RC Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL391041; CAC01575.1; -.
SQ SEQUENCE 156 AA; 17188 MW; 7181972EE7F33FB0 CRC64;

Query Match 21.3%; Score 150; DB 2; Length 156;
Best Local Similarity 34.7%; Pred. No. 3.6e-07;
Matches 34; Conservative 14; Mismatches 50; Indels 0; Gaps 0;

Qy 13 IPVELRYEGDPAIRMTFHLPGDAPVTWAFGRELLLDGLNSPGDGDVHIGTPESGLG 72
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 37 VPAEFYDDPHLLVLTIRFAPEGAPPVTHVHGRDLLHEGLRTTSGLDGVQVWADPTDRE 96
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 73 DVHIRLQVGADRALFRAGTAPLVAFLDRTDKLVPLQGE 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 97 TAWLQVNAHGDAIFSLPVELEWIDRTYLHVPAgte 134
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
Q9X7R1 PRELIMINARY; PRT; 142 AA.
AC Q9X7R1:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE REGULATOR.
GN SC5H1.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RC Oliver K., Harris D.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL049863; CAB42928.1; -.
SQ SEQUENCE 142 AA; 15741 MW; E24AA52C00AF40F1 CRC64;

Query Match 19.9%; Score 140; DB 2; Length 142;
```

Best Local Similarity 31.0%; Pred. No. 3.1e-06;  
Matches 40; Conservative 20; Mismatches 51; Indels 18; Gaps 5;

QY 12 RPPV--ELRYEVDPAIRMTFHLPGDAPVTWAFGRELLDGLNPSGDDVHIGTPE 69  
DB 21 RPPVLAHAYDAADPALTVPFSDHGRVLRWLTDRMVAEGLTRPVGVGDVRLRP-ESR 79

QY 70 GLGDVHIRLQVAD-----RALFRAGTAPLVAFLDRTDKLVLPGQEHITLGDGDFGNLE 121  
DB 80 GMDW-ELRIELLDGGRADGERHRAVVFVWAAVAEFLRETHAVVRGEEV-----RVD 132

QY 122 DALGRILAE 130  
DB 133 DFLAELTAE 141

RESULT 11  
Q9RKC9 PRELIMINARY; PRT; 126 AA.  
AC Q9RKC9;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)  
DE HYPOTHETICAL 13.7 KDA PROTEIN.  
GN SC87.09C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL132674; CAB59654.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 126 AA; 13742 MW; D54A8574D28B4D69 CRC64;

Query Match 14.1%; Score 99.5; DB 2; Length 126;  
Best Local Similarity 32.4%; Pred. No. 0.025;  
Matches 34; Conservative 10; Mismatches 30; Indels 31; Gaps 6;

QY 13 IPVELRYEVDPAIRMTFHLPG-----DAPVTWAFGRELLDGLNPSGDDVHIGTPT 66  
DB 27 IPVVLRYDLPDPSKVCVA--LPGRGRASGSRWTFSLLEOGLRAPAGSGEVRVWPC 84

QY 67 -----EPEGLGDVHIRLQVADRALFRAGTAPLVAFLDRT 101  
DB 85 GRVQAVFEHSPQGSVVQFE-----NKALIR-----FLRRT 116

RESULT 12  
Q9HSI1 PRELIMINARY; PRT; 548 AA.  
AC Q9HSI1;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)

DE TYPE II SECRETION SYSTEM PROTEIN.  
GN GSPE1 OR VNG0218G.  
OS Halobacterium sp. (strain NRC-1).  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OC Halobacterium.  
OX NCBI\_TaxID=64091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,  
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
RT "Genome sequence of Halobacterium species NRC-1."  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
DR EMBL; AE004986; AAG18825.1; -.  
DR InterPro; IPR001687; ATP\_GTP\_A.  
DR InterPro; IPR000619; Guanylate\_kin.  
DR PROSITE; PS0052; GUANYLATE\_KINASE\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 548 AA; 58788 MW; 7464274B1C7D9D2F CRC64;

Query Match 12.0%; Score 84.5; DB 17; Length 548;  
Best Local Similarity 26.6%; Pred. No. 4.7;  
Matches 41; Conservative 11; Mismatches 45; Indels 57; Gaps 7;

QY 11 FRIPVELRYEVDG-----PYAIRMTFHLPGDAPVTWAFGRELLDGLNPSGD-GDVH 62  
DB 184 FIVPVRIYTPGDADAFITTYAVREPDLPGDARIACKDRITNGVATGSDAGDDH 243

QY 63 IGTPEGLGDVHIRLQVADRA---LFRAGTAPLV-----AFLDRTDKLPV 106  
DB 244 AAAT-----ADRARSLLARRHTAPKTRHWCAGTADRVREAVADRGPAVPP 288

QY 107 LQGEHT-----LGDF-----DGNLED 122  
DB 289 VGDHVDLAVYVVRDLVGDAELTIPIRDPNLED 322

RESULT 13  
Q9XVD2 PRELIMINARY; PRT; 261 AA.  
AC Q9XVD2;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE C24H11.6 PROTEIN.  
GN C24H11.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lloyd C.R.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology."  
RL Science 282:2012-2018(1998).  
DR EMBL; Z81475; CAB03913.1; -.  
DR MEROPS; S26.013; -.  
DR InterPro; IPR000508; Peptidase\_S26.  
DR InterPro; IPR000223; Peptidase\_S26A.  
DR Pfam; PF00461; Peptidase\_S26; 1.





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:21:43 ; Search time 4085.76 Seconds  
(without alignments)  
2084.582 Million cell updates/sec

Title: US-09-749-185-4  
Perfect score: 407  
Sequence: 1 atgagcttcctcgtctcga.....gaggagcagagcgcggtg 407

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_em.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Score Match Length DB ID Description

1	407	100.0	407	6	AX007220	AX007220 Sequence
2	407	100.0	566	1	AF195771	AF195771 Streptomy
3	248.6	61.1	15441	1	SCQ11	AL096823 Streptomy
4	243.8	59.9	407	6	AX007222	AX007222 Sequence
5	243.8	59.9	408	1	AF195770	AF195770 Streptomy
6	239	58.7	438	6	AX007218	AX007218 Sequence
7	239	58.7	438	6	AX007216	AX007216 Sequence
8	239	58.7	1513	1	D50051	D50051 Streptomyce
9	224.6	55.2	407	6	AX007224	AX007224 Sequence
10	224.6	55.2	438	1	AF195772	AF195772 Streptomy
11	135.8	33.4	1410	1	STMLACBG	M28303 S.albus bet
12	110	27.0	38640	1	SCL2	AL137778 Streptomy
13	109.6	26.9	35284	1	SCE19A	AL096852 Streptomy
14	77.2	19.0	32704	1	SC8A11	AL391041 Streptomy
15	74.4	18.3	37245	1	SC5F2A	AL049587 Streptomy
16	73	17.9	45624	6	AX089419	AX089419 Sequence
17	73	17.9	50000	6	AX089416	AX089416 Sequence
18	67.8	16.7	42527	1	SC51A	AL121596 Streptomy
19	66.8	16.4	10417	1	AE004546	AE004546 Pseudomon
20	64	15.7	33285	1	SC5F8	AL357613 Streptomy
21	62.8	15.4	110000	2	LMFLCHR36_07	Continuation (8 of
22	62.6	15.4	173339	2	OSJN00060	AL606633 Oryza sat
23	62.2	15.3	30941	1	SC4C6	AL079355 Streptomy
24	61.2	15.0	110000	2	LMFLCHR32_06	Continuation (7 of
25	60.8	14.9	136775	2	AP004279	AP004279 Oryza sat
26	60	14.7	147452	8	AC026815	AC026815 Oryza sat
27	59.8	14.7	110000	2	LMFLCHR32_06	Continuation (7 of
28	59.6	14.6	113193	1	AF357202	AF357202 Streptomy
29	59.6	14.6	137174	8	AP002484	AP002484 Oryza sat
30	59.2	14.5	24700	1	SCE8	AL035654 Streptomy
31	59	14.5	20141	1	SC2A6	AL389898 Streptomy
32	58.6	14.4	10306	1	AE005712	AE005712 Caulobact
33	58.6	14.4	36368	1	SC9B5	AL035206 Streptomy
34	58.6	14.4	39531	1	SCP8	AL390975 Streptomy
35	58.4	14.3	3596	1	SAPKSGENE	Z46913 S.ambofacie
36	58.4	14.3	77457	1	AF210249	AF210249 Streptomy
37	58.2	14.3	8058	1	SCE56	AL138852 Streptomy
38	58.2	14.3	25458	1	SCE22	AL358332 Streptomy
39	58.2	14.3	37898	1	SCE34	AL353862 Streptomy
40	58.2	14.3	71286	1	AF143772	AF143772 Mycobacte
41	58	14.3	12970	1	SPU10405	U10405 Streptomyce
42	58	14.3	146836	8	AC074105	AC074105 Oryza sat
43	57.8	14.2	3867	1	AY033407	AY033407 Myxococcu
44	57.4	14.1	22115	1	SCC22	AL096839 Streptomy
45	57.2	14.1	2453	8	AY037870	AY037870 Oryza sat

ALIGNMENTS

RESULT 1	AX007220	AX007220	Sequence 5 from Patent WO0000613.	407 bp	DNA	linear	PAT 06-SEP-2000
LOCUS	AX007220	Sequence 5 from Patent WO0000613.					
DEFINITION	AX007220	AX007220.1	GI:9995086				
ACCESSION	AX007220	AX007220.1	GI:9995086				
VERSION	AX007220	AX007220.1	GI:9995086				
KEYWORDS	AX007220	AX007220.1	GI:9995086				
SOURCE	AX007220	AX007220.1	GI:9995086				
ORGANISM	AX007220	AX007220.1	GI:9995086				
REFERENCE	AX007220	AX007220.1	GI:9995086				
AUTHORS	AX007220	AX007220.1	GI:9995086				
TITLE	AX007220	AX007220.1	GI:9995086				
JOURNAL	AX007220	AX007220.1	GI:9995086				
FEATURES	AX007220	AX007220.1	GI:9995086				
source	AX007220	AX007220.1	GI:9995086				
Location/Qualifiers	AX007220	AX007220.1	GI:9995086				
1..407	AX007220	AX007220.1	GI:9995086				
/db_xref="taxon:1962"	AX007220	AX007220.1	GI:9995086				
Streptomyces albus G.	AX007220	AX007220.1	GI:9995086				
Streptomyces albus G.	AX007220	AX007220.1	GI:9995086				
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	AX007220	AX007220.1	GI:9995086				
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	AX007220	AX007220.1	GI:9995086				
1 (bases 1 to 407)	AX007220	AX007220.1	GI:9995086				
Kraal,B., Luiten,R.G. and Van Wezel,G.P.	AX007220	AX007220.1	GI:9995086				
Reducing branching and enhancing fragmentation in culturing	AX007220	AX007220.1	GI:9995086				
filamentous microorganisms	AX007220	AX007220.1	GI:9995086				
Patent: WO 000613-A 5 06-JAN-2000;	AX007220	AX007220.1	GI:9995086				
UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA	AX007220	AX007220.1	GI:9995086				
(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS	AX007220	AX007220.1	GI:9995086				
VAN (NL)	AX007220	AX007220.1	GI:9995086				



JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL

for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351  
2 (bases 1 to 15441)  
Seeger,K. and Harris,D.  
Unpublished  
3 (bases 1 to 15441)  
James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.  
Submitted (08-JUL-1999) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: Barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT

Streptomyces coelicolor sequencing at The Sanger Centre is funded  
by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
CDS are numbered using the following system eg SC7B7.01c, SC (S.  
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary  
strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons  
using a specially developed Hidden Markov Model (Krogh et al.,  
Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot  
program of Bibb et al., Gene 30:157-66(1984) as implemented at  
<http://www.nh.90.jp/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the  
correct initiation codon. Where possible we choose an initiation  
codon (atg, gtg, ttg or (att)) which is preceded by an upstream  
ribosome binding site sequence (optimally 5-13bp before the  
initiation codon). If this cannot be identified we choose the most  
upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions.  
Cosmid Q11 Cosmid Q11 lies between H24 and D78 on the AseI-Q  
genomic restriction fragment.

FEATURES  
source

Location/Qualifiers  
1. .15441  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid Q11"  
complement(1. .289)  
/gene="SCQ11.01c"  
complement(<1. .289)  
/gene="SCQ11.01c"  
/note="SCQ11.01c, hypothetical protein, partial CDS, len:  
>95 aa, unknown function, previously sequenced as  
TR:Q53870 (EMBL:U37580), S.coelicolor hypothetical protein  
ORF5 downstream of phosphotyrosine protein phosphatase  
(159 aa). Weak similarity to the N-terminus of TR:O69888  
(EMBL:AL023797) Streptomyces coelicolor hypothetical  
protein (172 aa) fasta scores; opt: 158 z-score: 206.4  
E(): 0.00036, 46.2% identity in 65 aa overlap. Continues  
as SCH24.40c in cosmid H24 (EMBL:AL049826)"  
/codon\_start=1  
/transl\_table=11  
/label="SCQ11.01c  
/product="hypothetical protein"  
/protein\_id="CAB46956.1"  
/db\_xref="GI:5457268"  
/translation="MTVRPVVKRARAVALDGDHILIKRTKPGVDPPWYTPGGGVEP

gene  
CDS

misc\_feature  
misc\_feature  
gene  
CDS

DDTTVDALHREVVEELGAKISDVVPCFVDVTEHIGEDGGATGKVKVQHFFVC"  
complement(1. .4279)  
/note="sequence corresponding to EMBL:U37580 from 1 to  
4279"  
1. .98  
/note="overlap with Streptomyces coelicolor cosmid H24  
(EMBL:AL049826) from 41528 to 41625"  
complement(360. .1265)  
/gene="SCQ11.02c"  
/note="abab"  
complement(360. .1265)  
/gene="SCQ11.02c"  
/note="SCQ11.02c, abab, probable lysR-family  
transcriptional regulator, len: 301 aa; previously  
sequenced as TR:Q53869 (EMBL:U37580) S.coelicolor  
lysR-like protein (301 aa). Highly similar to  
SW:ARAB\_STRAT (EMBL:X91393), araB, Streptomyces  
antibiotic transcriptional regulatory protein (301 aa),  
fasta scores; opt: 1581 z-score: 1813.2 E(): 0, 80.3%  
identity in 300 aa overlap. Similar to many other  
transcriptional regulators. Contains probable  
helix-turn-helix motif at aa 17-38 (Score 1659, +4.84 SD).  
Contains Pfam match to PF00126 HTH\_1, Bacterial regulatory  
helix-turn-helix protein, lysR family and PS00044  
Bacterial regulatory proteins, lysR family signature"  
/codon\_start=1  
/transl\_table=11  
/label="abab"  
/product="putative lysR-family transcriptional regulator"  
/protein\_id="CAB46957.1"  
/db\_xref="GI:5457269"  
/translation="MDIALLRFTVTVHRAGSFTRAAALGLSOPAVTSQIRTLERQLG  
RPLRLQARGVTPPTIGDELAKHAPHDALVEIAETACLEDSDTLRLHLGAPPEFTA  
ERALPGLGEGDQGFALRASFNGAETLEGLAAGHHDLAIGTRRAGHATPL  
CDEVHVLVATPHWAERAGVEDVDTASALKHVPVVEHSLPFGVRYWVSFFARPA  
SPATVAPDLRAVLACAVAGAGLAVLPRLYCAEALRGDVVALHDPVPVPLRTYFLV  
RTGTLAMPHIARAHEWLIRAAADWN"  
complement(831. .1256)  
/gene="SCQ11.02c"  
/note="Match to PF00126 HTH\_1, Bacterial regulatory  
helix-turn-helix protein, lysR family Score 145.96"  
complement(1125. .1217)  
/gene="SCQ11.02c"  
/note="PS00044 Bacterial regulatory proteins, lysR family  
signature"  
complement(1518. .2696)  
/gene="SCQ11.03c"  
/note="cysA"  
complement(1518. .2696)  
/gene="SCQ11.03c"  
/note="SCQ11.03c, cysA, probable cystathionine/methionine  
gamma-synthase/lyase, len: 392 aa; previously sequenced as  
SW:CYS\_STRCO (EMBL:U37580), cysA, S.coelicolor putative  
cystathionine gamma-lyase (392 aa). Similar to many e.g.  
SW:MEGL\_PSEPU (EMBL:D88554), mdeA, Pseudomonas putida  
methionine gamma-lyase (398 aa), fasta scores; opt: 608  
z-score: 653.6 E(): 4.5e-29, 36.9% identity in 404 aa  
overlap. Highly similar to TR:Q53668 (EMBL:X91393)  
Streptomyces antibioticus hypothetical protein found  
upstream of the abab gene (232 aa) (87.1% identity in 232  
aa overlap). Contains Pfam match to PF01053  
Cys Met Meta\_PP, Cys/Met metabolism PLP-dependent enzyme  
and PS00868 Cys/Met metabolism enzymes pyridoxal-phosphate  
attachment site"  
/codon\_start=1  
/transl\_table=11  
/label="cysA"  
/product="putative cystathionine/methionine  
gamma-synthase/lyase"  
/protein\_id="CAB46956.1"  
/db\_xref="GI:5457270"  
/translation="MSDSATTTDSAGTGGERSASAPGDDGTRAVRAGLPEPVKHEPTLPG  
PVFAAHPLPGDPTGPTVYGRDENTWTRLESAIGEAEAGVETLVFASGMAAIS

SVLFSQLRAGTAVLPDDGYQALPLVRAQLAEYIEVTRAPTGRDQALDVLGAKLLM  
ITFSPNGLDVRRLVEAAHAGGALVAVDNTLATPLGQPLELGADEFVASTKQL  
TGHGVDLLGVAGDRAAGVAARRWKIVGAIIPGMEANLAHRSIATLQURVDRQST  
ALKVAEALRPTETGLRPLCPDDPSHKVASOQMLRYGCWVFTLPSRARADREFLDA  
LRLVEGATSGGVRSRAERGRWGGDAVPEGFIRLSVGAEDPDDLVLADLLALDEYTE  
" complement(1533..2624)  
/gene="SCQ11.03c"  
metabolism Ptp-dependent enzyme Score 268.52"  
complement(2031..2075)  
/gene="PSO0868 Cys/Met metabolism enzymes  
pyridoxal-phosphate attachment site"  
complement(2693..3187)  
/gene="SCQ11.04c"  
/note="ptpA"  
complement(2693..3187)  
/gene="SCQ11.04c"  
/note="SCQ11.04c", ptpA, low molecular weight  
protein-tyrosine-phosphatase, len: 164 aa; previously  
sequenced as SW:PTPA\_STRCO (EMBL:U37580), ptpA,  
S.coelicolor low molecular weight  
protein-tyrosine-phosphatase (164 aa). Contains Pfam match  
to PF01451 LMWPC. Low molecular weight phosphotyrosine  
protein phosphatase"  
/codon\_start=-1  
/transl\_table=11  
/label=ptpA  
/product="low molecular weight  
protein-tyrosine-phosphatase"  
/protein\_id="CAB46959.1"  
/db\_xref="GI:5457271"  
/translation="MYRVCFVCTGNTICRSPMAEAFVAREVDAGLHLEADSCTG  
GWHEGEGADPTEAVLADHGYGLDHAARQFQSWFSLDLVVALDAGHLRALRLAPT  
ERDAKVRLLRSRDPVAVAGDLDPDPFYGGROGFECELEWEEAASGLLIAVREQVE  
GRAA"

Query Match 61.1%; Score 248.6; DB 1; Length 15441;  
Best Local Similarity 75.7%; Pred. No. 1.6e-19;  
Matches 308; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 1 atgagttctctgcgtccgagagagctcgcctccgcagatcccggtggagctgcgtacgag 60  
DB 5475 ATGAGCTTCTCGTCCGAGGAGCTCTCTTCGCGCATCCGCGTGAGCTCGCTACGAG 5534  
QY 61 accgtcagtcgtagcgggtcgcgtacgttccacctcccgcagagaccccggtcacc 120  
DB 5535 ACCCGGATCCCTATGCGGTACGCTGACCTTTCATCTGCCCGGAGACGCGCGGTGACC 5594  
QY 121 tgggtcttcgggctgaactgctggtcagggagctccttgacgcgcgcgcgcgcgcgcgc 180  
DB 5595 TGGGCTTCGCGCGGAGCTGCTGCTGACGGGTGGCGCGCGCTGCGCGGACGCGGAC 5654  
QY 181 gtcggggtctgcgggtggggcagacggccacacaggaggtgcacatcacctccaggtc 240  
DB 5655 GTGCGCATCGCGCGGTGGAGACCGGAGCGCGTGTGATGATGATGATGATGATGATGAT 5714  
QY 241 ggtccgagcagcgctcttcgcgtcgcgaaggcgcgcgcgcgcgcgcgcgcgcgcgcgc 300  
DB 5715 GGCAGCACACAGGCGCTGTCCTGCTCCCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 5774  
QY 301 accgaccagggcttgcgtcgc 360  
DB 5775 ACCGACAGCTGTGCTGCGCGAGAGGCGCGCTGCGCGGACCTTCGACAGCCACTC 5834  
QY 361 gacgagctctgaaccagcagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 407  
DB 5835 GACGAGCCCTGGACCGCATCTCTGGCGGAGACAGAGAGCGCGGCTG 5881

RESULT 4  
AX007222

AX007222 407 bp DNA linear PAT 06-SEP-2000  
Sequence 7 from Patent WO0000613.  
AX007222  
Version AX007222.1 GI:9995088  
KEYWORDS Streptomyces goldenensis.  
ORGANISM Streptomyces goldenensis  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 407)  
AUTHORS Kraal,B., Luiten,R.G. and Van Wezel,G.P.  
TITLE Reducing branching and enhancing fragmentation in culturing  
filamentous microorganisms  
JOURNAL Patent: WO 0000613-A 7 06-JAN-2000;  
UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJBERTUS MARIA  
(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS  
VAN (NL)  
FEATURES Location/Qualifiers  
source 1..407  
/organism="Streptomyces goldenensis"  
/db\_xref="taxon:121022"  
gene 1..405  
/gene="ssgA"  
CDS 1..>405  
/gene="ssgA"  
/note="strain ATCC of Streptomyces goldenensis"  
/codon\_start=1  
/transl\_table=11  
/product="SsgA"  
/protein\_id="CAC07387.1"  
/db\_xref="GI:9995089"  
/translation="MSFLVSEELSFRIPELVRYETCDPVAVRLTFHLPDAPVTWAFG  
RELLIDGPRPCGDGVHIAPADPETFGEVLIRLQVSDQAMFRVGTAPLVAFLDRDT  
KIVPLGGERSLADFALDLDEALDRILAEQNAQ"  
BASE COUNT 56 a 135 c 145 g 71 t  
ORIGIN

Query Match 59.9%; Score 243.8; DB 6; Length 407;  
Best Local Similarity 74.9%; Pred. No. 1.7e-18;  
Matches 305; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
QY 1 atgagttctctgcgtccgagagagctcgcctccgcagatcccggtggagctgcgtacgag 60  
DB 1 ATGAGCTTCTCGTCCGAGGAGCTCTCTTCGCTATTCGCGTGAGCTCGCTACGAG 60  
QY 61 accgtcagtcgtagcgggtcgcgtacgttccacctcccgcagagaccccggtcacc 120  
DB 61 ACCTGTATCCCTACGCGGTGCGGCTGACCTTTCATCTGCCCGGAGATGCCCGGTGACC 120  
QY 121 tgggtcttcgggctgaactgctggtcagggagctccttgacgcgcgcgcgcgcgcgcgc 180  
DB 121 TGGGCTTCGCGCGGAGTTCCTCATCGACGAGTCCGCGCGCTGCGGAGCGGAGAC 180  
QY 181 gtcggggtctgcgggtggggcagacggccacacaggaggtgcacatcacctccaggtc 240  
DB 181 GTCCACATCGCGCGCGCGACCGGAGACGTTCCGCGAGGTCCTGTATGATGATGATGAT 240  
QY 241 ggtccgagcagcgctcttcgcgtcgcgaaggcgcgcgcgcgcgcgcgcgcgcgcgcgc 300  
DB 241 GCGAGCGACGAGGAGTTCCTGCGGTGCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCT 300  
QY 301 accgaccagggcttgcgtcgc 360  
DB 301 ACGGACAAAGATCGTGGCGCTGGGCGAGAGGCTTCCTTCGCCGACTTCGACGCCCTC 360  
QY 361 gacgagctctgaaccagcagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 407  
DB 361 GACGAGCGCTGGACCGCATCTCTGGCGGAGACAGAGAGCGCGGCTG 407

RESULT 5  
AF195770



```
RESULT 7
AX007216
LOCUS AX007216 438 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0000613.
ACCESSION AX007216
VERSION AX007216.1 GI:9995082
KEYWORDS Streptomyces griseus.
SOURCE Streptomyces griseus.
ORGANISM Streptomyces griseus.
REFERENCE
AUTHORS Kraal,B., Luiten,R.G. and Van Wezel,G.P.
TITLE Reducing branching and enhancing fragmentation in culturing filamentous microorganisms
JOURNAL Patent: WO 000613-A 1 06-JAN-2000;
UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GJJSBERTUS MARIA (NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS VAN (NL)
FEATURES
SOURCE 1..438
/organism="Streptomyces griseus"
/db_xref="taxon:1911"
gene 1..408
/gene="ssgA"
CDS 1..>408
/gene="ssgA"
/note="strain ATCC of Streptomyces griseus"
/codon_start=1
/transl_table=11
/product="ssgA"
/db_xref="GI:9995083"
/translation="MRESVQAEVMSFLVSELSFRIPVELRYEVDPAIRMTFHLPGDAPVTFWAGRELLLDGLNSPGSDGVHIGTEPEGLGVHRLQVGADRALFRAGTA PLVAFVLDRLDKVLPGQHTLGDGFDGNLEALGR"
BASE COUNT 68 a 145 c 153 g 72 t
ORIGIN
Query Match 58.7%; Score 239; DB 6; Length 438;
Best Local Similarity 74.2%; Pred. No. 5.7e-18;
Matches 302; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY 1 atgagctctcgtctccgaggagctcgcttcgcatccgagtcgagctgcggtacgag 60
Db 31 ATGAGCTTCTCGTCTCCGAGGAGCTCTCGTTCGTRATCCGGTGGAGCTCCGATACGAG 90
QY 61 accgtcgatccgtacgcggtgcggtgcggtgcggtgcggtgcggtgcggtgcggt 120
Db 91 GTCGGCATCCGATGCGATCCGGATGACCTTCCACCTCCCGCGATGCCCTGTGACC 150
QY 121 tgggtcttcgggcgtgaactcgtggtcagaggagtcctgcagccgcgcgcgcgcgcgc 180
Db 151 TGGCGGTTCGGCCGCGAGCTGCTGCTGAGCGGGCTCAACAGCCGCGAGCGCGCGAT 210
QY 181 gtccgggtctgcccgtggtggcagacgcccaccaggaggtgcacatcacctccaggtc 240
Db 211 GTGCACATCGCGCCGACCGAGCGGAGGGCTCGGAGATGTCACATPCCGGCTCCAGTC 270
QY 241 ggctccgagcaggcgctcttcgctgcggcaaggcgcgctgctgccttccttcgaccgc 300
Db 271 GGC CGGACCGTGGCTGTTCGGGGCGGGGACGCGACCGCTGTTGGGTTCCTCGACCG 330
QY 301 accgaccagggtctgtcgtcgcagcagcggcgacacgcgcgcgcgcgcgcgcgcgcgc 360
Db 331 ACGGACAGTCTGCGCGCTCGGCGAGGACACACGCTGGGTGACTTCGACGCGCACTG 390
QY 361 gacgacgtctgaaccgcacgctcgcgagagagagagagagagagagagagagagagag 407
Db 391 GAGGACGCACTGGCGCCGATCTTCGCCGAGGAGCAGAACGCCGGCTG 437
```

```
RESULT 8
D50051
LOCUS D50051 1513 bp DNA linear BCT 19-MAY-1999
DEFINITION Streptomyces griseus DNA for ssgA, complete cds.
ACCESSION D50051
VERSION D50051.1 GI:1772323
KEYWORDS ssgA.
SOURCE Streptomyces griseus (strain:B2682) DNA.
ORGANISM Streptomyces griseus
REFERENCE
AUTHORS Kawamoto,S. and Ensign,J.C.
TITLE Cloning and characterization of a gene involved in regulation of sporulation and cell division of Streptomyces griseus
JOURNAL Actinomycetol. 9, 136-151 (1995)
REFERENCE
AUTHORS Kawamoto,S. and Ensign,J.C.
TITLE Isolation of mutants of Streptomyces griseus that sporulate in nutrient rich media
JOURNAL Actinomycetol. 9, 124-135 (1995)
REFERENCE
AUTHORS Kawamoto,S., Watanabe,H., Hesketh,A., Ensign,J.C. and Ochi,K.
TITLE Expression analysis of the ssgA gene product, associated with sporulation and cell division in Streptomyces griseus
JOURNAL Microbiology (Reading, England) 143 (Pt 4), 1077-1086 (1997)
MEDLINE 97286526
PUBMED 9141673
REFERENCE
AUTHORS Shinichi,K. and Ensign,J.
TITLE Cloning and characterization of a gene involved in sporulation and cell division of Streptomyces griseus
JOURNAL Unpublished (1995)
REFERENCE
AUTHORS Shinichi,K.
TITLE Direct Submission
JOURNAL Submitted (06-APR-1995) Kawamoto Shinichi, National Food Research Institute, Biomolecular Transfunction Laboratory; Kannondai 2-1-2, Tsukuba, Ibaraki 305, Japan (Tel:0298-38-8124, Fax:0298-38-7996)
FEATURES
SOURCE 1..1513
/organism="Streptomyces griseus"
/strain="B2682"
/db_xref="taxon:1911"
RBS 385..389
CDS 392..802
/codon_start=1
/transl_table=11
/product="ssgA"
/protein_id="BAA21558.1"
/db_xref="GI:2281004"
/translation="MMSFLVSELSFRIPVELRYEVDPAIRMTFHLPGDAPVTFWAGRELLLDGLNSPGSDGVHIGTEPEGLGVHRLQVGADRALFRAGTAPLVAFDRTDKLVPGLQHTLGDGFDGNLEALGRILAEENAG"
stem_loop 839..875
stem_loop 863..968
BASE COUNT 251 a 539 c 493 g 230 t
ORIGIN
Query Match 58.7%; Score 239; DB 1; Length 1513;
Best Local Similarity 74.2%; Pred. No. 3.9e-18;
Matches 302; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY 1 atgagctctcgtctccgaggagctcgcttcgcatccgagtcgagctgcggtacgag 60
Db 395 ATGAGCTTCTCGTCTCCGAGGAGCTCTCGTTCGTRATCCGGTGGAGCTCCGATACGAG 454
QY 61 accgtcgatccgtacgcggtgcggtgcggtgcggtgcggtgcggtgcggtgcggtgcg 120
Db 455 GTCGGCATCCGATGCGATCCCGATCCCGATCCCGATCCCGATCCCGATCCCGATCCCG 514
```

Qy	121	tgggtctctcgggcgcgaactgctggtcgagggagtcctggaagccgagcgagcgagc	180
Db	515	TGGGCGTTCGGCCGCGAGCTGCTGCTGGACGGGCTCAACAGCCGAGCGGCGAGCGGGAT	574
Qy	181	gtccgggtctgcgccgtggtgggcagacgcccacaggaggtgcacatcacctccaggtc	240
Db	575	GTGCACATCGCCCGGACCGAGCCGAGGGCTCGGAGATGTCCACATTCGGCTCCAGGTC	634
Qy	241	ggctccgagcaggcgtctcttcgcgtcggcaagcgccgctgtcgtctcttcgcagcgc	300
Db	635	GGCGCGAGCCGTGCGCTGTTCGGCGCGGGACGACACCGCTGTGTGGGTCTCTCGACGG	694
Qy	301	accgaccagggcttgtcgtctcgagcagcggggcacacccgactctcgacagccactc	360
Db	695	ACGGACAAGCTCGTGCCTCGCTCGCCAGGAGCACACGCTGGGTGACTTCGAGGCAACTG	754
Qy	361	gacgacgctctgaacgcgagccttcgcgcgagggagcagagccgcggctg	407
Db	755	GAGAGCACTGGCGCCGCATCTCTCGCGAGGAGCAGAACCGCGGCTG	801

RESULT	9	
AX007224		
LOCUS	AX007224	
DEFINITION	Sequence 9 from Patent WO0000613.	407 bp DNA linear PAT 06-SEP-2000
ACCESSION	AX007224	
VERSION	AX007224.1	GI:9995090
KEYWORDS		
SOURCE	Streptomyces netropsis.	
ORGANISM	Streptomyces netropsis	
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE	1 (bases 1 to 407)	
AUTHORS	Kraal,B., Luiten,R.G. and Van Wezel,G.P.	
TITLE	Reducing branching and enhancing fragmentation in culturing filamentous microorganisms	
JOURNAL	Patent: WO 0000613-A 9 06-JAN-2000;	
	UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA	
	(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS	
VAN (NL)		
FEATURES	Location/Qualifiers	
source	1..407	/organism="Streptomyces netropsis"
		/db_xref="taxon:55404"
gene	1..405	
		/gene="ssgA"
CDS	1..>405	
		/gene="ssgA"
		/note="strain ATCC of Streptomyces netropsis"
		/codon_start=1
		/transl_table=11
		/product="SsgA"
		/protein_id="CAC07388.1"
		/db_xref="GI:9995091"
		/translation="MSFLSEELSFKIVELRYETRDPAVRMTFHLPGDAPVTFWAFG
		RESLDGIINRPSGDGVHAPTDPEGLSDVSRILQVADRALFRAGAPPLVFLDRFT
		KSVPLGQETLGDGFDSLEAALGKILAEQNAG"
BASE COUNT	63 a 151 c 131 g 62 t	
ORIGIN		

		Query Match	55.2%	Score 224.6;	DB 6;	Length 407;
		Best Local Similarity	72.0%;	Pred. No. 2.3e-16;		
		Matches 293;	Conservative 0;	Mismatches 114;	Indels 0;	Gaps 0;
QY	1	atagcttcctcgtctccgaggagctgccttcgcgatccccggtggagctgcggtacag	60			
Db	1	ATGAGCTTCCGTCTCCGAGGAGCTCCTTCAAGATCCGACTCGAATCGCATACGAG	60			
QY	61	accgtcgatccgtacgcggtgcggctgaagtccacctcccggagagacgcccggtcaccc	120			
Db	61	ACCGGGATCCCTACCGGGTCGGATGACCTTCACCTCCCGSAGACGGCGCTGTGACC	120			

Qy	121	tggtctcttcggcgctgaactgctggtcgagggagtctctgacgcgcggcgacgagcgac	180
Db	121	TGGCGTTCGGCGGGAGCTCTGCTGCACGGGATCAACCGCCGAGCGCGGACGGCGAC	180
Qy	181	gtcggggtctaccgggtgggcagacagccaccagggaggtgcacatcacctccagctc	240
Db	181	GTCCACATCGCCCGACCGCAGGGCTGTTCGACGTCTCCATCCGGCTCCAGGTG	240
Qy	241	ggctccgagcaggcgctcttccgcgctgcgcgaagcgccgcgtctgcgcttctcgacgc	300
Db	241	GGCGGGACCGCGGCCCTCTCCGTGCAGGGCGCCCGCGCTGTGTCGCTTCTCTGACCGC	300
Qy	301	accacacagggtctgcgtcgagcagagcggcacacccgacttcgcagaccactc	360
Db	301	ACGGACATCGTGTCCCTCTGGTCAGGAACAGACTCTGGTGACTTCGAGGACAGCCTG	360
Qy	361	gacgacgctctgaaccgcgactcgcgcgagagcagagcgcggctg	407
Db	361	GAGGCCCGGCTTCGGCAAGATCTCCGCGAGCAGCAGAACGCCGGCTG	407

RESULT	10
AF195772	
LOCUS	438 bp DNA linear BCT 01-NOV-2000
DEFINITION	Streptomyces netropsis SsgA (ssgA) gene, complete cds.
ACCESSION	AF195772
VERSION	AF195772.1 GI:11066162
KEYWORDS	.
SOURCE	Streptomyces netropsis.
ORGANISM	Streptomyces netropsis Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 438)
REFERENCE	van Wezel,G.P., Rousseau,C. and Kraal,B. Cloning and sequencing of the Streptomyces netropsis ssgA gene
AUTHORS	Unpublished
JOURNAL	2 (bases 1 to 438)
REFERENCE	van Wezel,G.P.
AUTHORS	Direct Submission
TITLE	Submitted (18-OCT-1999) Biochemistry, Unlversity of Leiden, PO Box
JOURNAL	9502, Leiden 2300 RA, Netherlands
FEATURES	Location/Qualifiers
source	1..438 /organism="Streptomyces netropsis" /db_xref="taxon:55404"

```

/gene="ssgA"
/note="similar to Streptomyces griseus ssgA; possible cell
division protein"
/codon_start=1
/transl_table=11
/product="SsgA"
/protein_id="FAG28483.1"
/db_xref="GI:11066163"
/translation="MRESVQAEVLMSELSFKIPVELRYETRDPTAVRMTFPLP
DPVPTWAFGRGELLGGINPSGDGDVHIAPTDPEGLSDYSIRLQVGADRALFRAGAP
PLVAFIDRTKSPVLCQEQTLDGDFDPSLEAALGKILAEQNAQ"
70 a 158 c 142 g 68 t
BASE COUNT
ORIGIN

```

	Query Match	55.2%	Score 224..6;	DB 1;	Length 438;
	Best Local Similarity	72.0%	Pred. No. 2.2e-16;		
	Matches 293;	Conservative	0;	Mismatches 114;	Indels 0; Gaps 0;
Qy	1	atgagctcttcgtctccgagagagctcgcccttcgcgcatccggtggagctgcggtacgag	60		
Db	31	ATGAGCTCTTCGTCGAGAGAGCTCTCTTCAAGATCGCAGTCGATACGAG	90		
Oy	61	acctctcatccgacgaggtacgctgaagcttccacactccccagagagcgcgcggtacc	120		

```
Db 91 ACCCGGATCCTACGGGTGCGATGACCTTCACCTCCCGGAGACGGCCCTGTGACC 150
Qy 121 tgggtcttcgggctgaactctggtcgagggagctcctgaacgcgcgagcgacgac 180
Db 151 TGGCGGTTCGCCGGGAGCTCTCTCTCAGCGGATCAACCGCCCGAGCGGCGAC 210
Qy 181 gtccgggtctgcccgggtgggagacagcgccaccaggagggtgcacatcacctccaggtc 240
Db 211 GTCCACATCGCCCGACCGACCGGCTGTGTCGACGCTCTCATCCGCTCCAGGTG 270
Qy 241 ggctccagagcaggcgtcttcctcgctcggaagcgccgctgctgccttctcgcagcgc 300
Db 271 GGC CGCGACCGCGCCCTCTTCCTGTCAGAGCGCCCGCGCTGCTGCGCTTCTCGACCGC 330
Qy 301 accgacagggtctgctcgtcgagcagcgagcgccacacgagcgccacgacttcgacgacctc 360
Db 331 ACGGACAGTCCGTCGCCCTGTCAGAACAGACTCTGGTGACTTCGAGGACAGCCTG 390
Qy 361 gacgacgtctgaaccgacgctcgcgagagcagagcgccggtcg 407
Db 391 GAGCGCGCTCGCGAAGATCTCGCGAGGAGCAGACCGCGCTG 437

RESULT 11
STMLACBG
LOCUS
DEFINITION S.albus beta-lactamase gene, complete cds.
ACCESSION M28303
VERSION M28303.1 GI:153338
KEYWORDS beta-lactamase.
SOURCE Streptomyces albus (strain G) DNA, clone pDML6.
ORGANISM Streptomyces albus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 1410)
AUTHORS Dehotay,P., Dusart,J., De Meester,F., Joris,B., Van Beeumen,J.,
Epicum,T., Frere,J.-M. and Ghuyssen,J.-M.
TITLE Nucleotide sequence of the gene encoding the Streptomyces albus G
beta-lactamase precursor
JOURNAL Eur. J. Biochem. 166, 345-350 (1987)
MEDLINE 87275916
FEATURES
    source          Location/Qualifiers
                    1..1410
                    /organism="Streptomyces albus"
                    /strain="G"
                    /db_xref="taxon:1888"
    gene            243..1242
                    /gene="beta-lactamase"
    CDS              243..1187
                    /gene="beta-lactamase"
                    /codon_start=1
                    /transl_table=11
                    /product="beta-lactamase"
                    /protein_id="AAA26775.1"
                    /db_xref="GI:153339"
                    /translation="MHPSTRSRPRTLLTATAGAAALAAATLVPGTAAHSSGGRGHSGG
SVSDARRLAGLRASARGLVAYDTGSRVAYRDELFPMSVEKTLSSAALRD
LDNRGFLSRILYTODDEQADGAPETGKPFONLANAQLTVEELCEVSITASDNC
NLMRLGGPAAVTRVSLGDRVTRLDWRPELNSAEPGRVDTTSPRAITRTYGR
VLGDLNPRDLRLTSLMANTTSGDRFRAGLPDDWTGLGKTGAGRYGTNNDAGVTWP
PGRAPVLTVLTAKEQDAARDGLVADAARVLAETLG"
    repeat_unit     1204..1216
                    /gene="beta-lactamase"
                    /note="inverted repeat"
    repeat_unit     1230..1242
                    /gene="beta-lactamase"
                    /note="inverted repeat"
BASE COUNT 196 a 540 c 496 g 178 t
ORIGIN
```

Query Match 33.4%; Score 135.8; DB 1; Length 1410;

```
Best Local Similarity 98.6%; Pred. No. 9e-07;
Matches 137; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 22 gactcgccttcacatcccggtgagctcgggtacgacacgcgacggtcgcgtacggtg 81
Db 1 GAGTCGCTTCGCATCCCGGTGGAGCTGGGTACGAGACCGTCGATCCGTACGCCGTG 60
Qy 82 'cgggtcaggttccacatcccgagacgcccgggtcaccctgggtcttcggggtgaaactg 141
Db 61 CGGTGACGTTCACCTCCCGGAGACGCCCGGTACCTGGGTCTTCGGGCGTGAACGTG 120
Qy 142 ctggtcagggagtcctcgtg 160
Db 121 CTGTCGAGGGAGTCAAGG 139

RESULT 12
SCL2
LOCUS
DEFINITION Streptomyces coelicolor cosmid L2.
ACCESSION AL137778
VERSION AL137778.1 GI:6822206
KEYWORDS acetyltransferase; acyltransferase; adenine
phosphoribosyltransferase; DNA helicase; DNA polymerase III;
elongation factor; endodeoxyribonuclease; GTP pyrophosphokinase;
membrane protein; membrane transferase; protein-export membrane
protein; regulator; secreted protein; sugar transferase;
threonine-tRNA synthetase; transport system inner membrane protein.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 38640)
AUTHORS Redenbach,M., Kieser,H.M., Denapaita,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
REFERENCE 2 (bases 1 to 38640)
AUTHORS Seeger,K.J. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 38640)
AUTHORS Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
Where these have been used to deduce the initiation codon.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.go.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
```





Query Match 27.08; Score 110; DB 1; Length 38640;  
Best Local Similarity 55.1%; Pred. No. 0.00022;  
Matches 215; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 2 tgagcttctcgttcgcgaggagctcgcttccgcatcccggtagctgcgggtacgaga 61  
Db 35030 TGC GCCCTGTTGTTCAGCGAGTCTCCCTGCCCTGTCCCCGACGGCTGCGGTACGACA 35089

QY 62 ccgtcgatccgtacgcggtgcggtgcgcttccaccctcccggagagccccgccct 121  
Db 35090 CGGCCGACCCTACGCGTGTCACGCCACTTCCACACCGGAGCGGAGAGACCGTCGAGT 35149

QY 122 gggcttcggggctgaactgctgtagggagtcttgagacgcgcggcagcgcgacg 181  
Db 35150 GG GTTTCGCCCGGCACTGCTCGCCGAAGTCTCCACCGCTCCACCGCACCGCGCAGC 35209

QY 182 tccgggtctcccggtgggcagacgcccaccagggaggtcacatcacctccaggtcg 241  
Db 35210 TCCGTGTCTGCCCTGCCGAGTCAGGCCAGGCGCTGTGTGATCGCTCTCAGCTCCC 35269

QY 242 gtccgagcagcgctcttccgcgtcgccaaggcgccgctgctgccttctccaccga 301  
Db 35270 CGGAGGCGGAGGCACTGCTCGAGGCCCGCGCGCGCCCTGGAGTCTTCTCGAAGCCA 35329

QY 302 ccgaccagggtttcgtcgtggcagcagcgaggggcacacgccagattgacagccacctcg 361  
Db 35330 CAGACGCGCGCTGCCCGCCGCGAACACCGGCACCTTCGATTCGATCAGGAGCTCT 35389

QY 362 acgagctctgaaccgcagctcgccagg 391  
Db 35390 CGCATCTCTGGCGGAAGCTTAGGCGCGGG 35419

RESULT 13

SCE19A  
LOCUS SCE19A linear BCT 16-JUL-1999  
DEFINITION Streptomyces coelicolor cosmid E19A.  
ACCESSION AL096852  
VERSION AL096852.1 GI:5531349  
KEYWORDS 4-hydroxyphenylpyruvate dioxygenase; ABC transporter; amino acid permease; asnC-family; cysteine synthase; nicotinamide; osmoregulation; osmotolerance; oxidoreductase; permealase; phosphoenolpyruvate-dependent; protease; regulator; ribonuclease PH; rph; sugar phosphotransferase system; transcriptional regulator; transposase; tRNA Leu; zinc metalloproteinase.  
SOURCE Streptomyces coelicolor A3(2).  
ORGANISM Streptomyces coelicolor A3(2).  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 35284)  
Redenbach,M., Kieser,H.M., Denapait,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.  
A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351  
2 (bases 1 to 35284)  
Seeger,K. and Harris,D.  
Unpublished  
3 (bases 1 to 35284)  
James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.  
Submitted (16-JUL-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK  
Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: <http://www.sanger.ac.uk/projects/S-coelicolor/>)



JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351  
2 (bases 1 to 32704)  
Saunders,D.C. and Harris,D.  
Unpublished  
3 (bases 1 to 32704)  
Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.  
Submitted (03-AUG-2000) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/s\\_coelicolor/](http://www.sanger.ac.uk/Projects/s_coelicolor/))  
CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (Optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid 8A11.

FEATURES  
SOURCE

Location/Qualifiers  
1..32704  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 8A11"

misc\_feature

1..100  
/gene="SC8A11.01"  
/note="nominal overlap with Streptomyces coelicolor cosmid St9A4"

gene

1..2254

CDS

/gene="SC8A11.01"

<1..2254  
/note="SC8A11.01"  
(fragment), len: >750 aa; similar to TR:Q9S124 (EMBL:AL109747) Streptomyces coelicolor probable transcriptional regulator SCU21.13, 919 aa; fasta scores: opt: 818 z-score: 913.7 E(): 0; 32.6% identity in 775 aa overlap. Contains Pfam match to entry PF00196 GerE, Bacterial regulatory proteins, luxR family and match to Prosite entry PS00622 Bacterial regulatory proteins, luxR family signature. Also contains a possible helix-turn-helix motif at residues 699..720 (+3.25 SD)"  
/codon\_start=2  
/transl\_table=11  
/product="putative transcriptional regulator (fragment)"  
/protein\_id="CAC01573.1"

/db\_xref="GI:9716212"  
/translation="TCAARTSPGTYTGLTHAELVLEPLSPSAAAEELLDADAPGLDA  
VRARILQRAAGNPLALVELPRAAQGISPLDLDLPLTORLETAFASTDSLTRCRTFL  
LVLAAPETAPLNQLLDVASLAGSEVTVVALQEAADVAGLVVLTRGTPRHPLMRSAI  
YTRAVVADRLSTHRALEETLEGSPGRVHLAAATGLGPODELAGOLERFADDAQKRGQ  
LAAAPALQAGELVHDPRRQGLLVRAAEALASEINDRVQAOILLNRADIAEPGPTER  
ARLMVSDKAAEFDEPEPORIDMDIAAAGAFDVGVSTVAENLLRRAACRFFQDGA  
RYRAQAAAEEDRWKPPDPAPHVLTTRAYTEPYRRGTDLJLARKLEKRPREDGRLHLL  
GSLAAGTGVGTRATRYLAQAAWSRQSLGALLARSLAGSWPRLYLQOLAQAARESAE  
GTGAETGEMVWLGLKATSAITAVLGEREAAASVRELRAHSLFFVPMPSASMAQ  
QVEGLLALFDSRAVEAYDALARAFDKTDPHYHSTSRWLLVDPDLYDAAAAGNEQARE  
LIVLEPELADRLPSEMIVARTYSTAVLPDDTDAEDCYDSALSALPDTWPLARAUHL  
QHRRLLRQRNRDARKLRLARDEFDRVGAQPMADMEQRAAGESQRRRPSKGE  
SLTVQEROIAELASQGLSNREIQGLRFSHRTVGAHLRYIYPRLGITTSRGLSAAALA  
LGDDQPTSGS"  
2042..2239  
/gene="SC8A11.01"  
/note="Pfam match to entry PF00196 GerE, Bacterial  
regulatory proteins, luxR family, score 49.10, E-value  
9.7e-11"  
2093..2176  
/gene="SC8A11.01"  
/note="PS00622 Bacterial regulatory proteins, luxR family  
signature"  
complement(2184..2453)  
/gene="SC8A11.02c"  
complement(2184..2453)  
/gene="SC8A11.02c"  
/note="SC8A11.02c, unknown, len: 89 aa"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC8A11.02c"  
/protein\_id="CAC01574.1"  
/db\_xref="GI:9716213"  
/translation="MASAAATAGTAGALPVMGQAHADTRGLPSLDWDVRSVVNPENA  
YTVTVQVNAEVAQLMGATFEETPRSRAGLRAPAPRRACRGW"  
2787..3257  
/gene="SC8A11.03"  
2787..3257  
/gene="SC8A11.03"  
/note="SC8A11.03, possible regulator, len: 156 aa; similar  
to TR:Q9S2F7 (EMBL:AL096852) Streptomyces coelicolor  
putative regulator SC819A.24, 142 aa; fasta scores: opt:  
272 z-score: 353.0 E(): 3.4e-12; 36.9% identity in 130 aa  
overlap"  
/codon\_start=1  
/transl\_table=11  
/product="putative regulator"  
/protein\_id="CAC01575.1"  
/db\_xref="GI:9716214"  
/translation="MSGDHHGVQAQHSAAQALLPLSLCSQMTGALEWEDVPAEFYRD  
PDHLLVTRFAPEGAPPVTHVGRDLLHEGLRTTSGLDGVQVWADTPTDRETAMLVQV  
NAHGDIATFSLPVPELEWIDRTYLHVPAAGTSSRLGTDAFLSKLDFDEPEASSR"  
complement(3297..7058)  
/gene="SC8A11.04c"  
complement(3297..7058)  
/note="SC8A11.04c, probable secreted peptidase, len: 1253  
aa; similar to TR:Q9RL54 (EMBL:AL121596) Streptomyces  
coelicolor probable secreted peptidase SCF51A.10, 1245 aa;  
fasta scores: opt: 1827 z-score: 1848.3 E(): 0; 37.2%  
identity in 1300 aa overlap, to TR:P95684 (EMBL:D83672)  
Streptomyces albogriseolus subtilisin-like protease, 1102  
aa; fasta scores: opt: 1365 z-score: 1380.6 E(): 0; 33.4%  
identity in 1268 aa overlap and to Streptomyces coelicolor  
probable secreted peptidase SC8A11.16c, 1239 aa; fasta  
scores: opt: 1835 z-score: 1467.4 E(): 0; 46.9% identity  
in 1279 aa overlap. Contains Pfam matches to entries  
PF00082 Peptidase S8, Subtilase family and PF02225 PA, PA  
domain and matches to Prosite entries PS00136 Serine  
proteases, subtilase family, aspartic acid active site,  
PS00137 Serine proteases, subtilase family, histidine  
active site and PS00138 Serine proteases, subtilase





Db 6006 CCAGGCCGCCGCCGAGCTGGTCCCGTGGGTCTCGAGCACCTCCAGCTGGACCTGGACCA 5947  
Qy 354 ccacctcgacgacgctctgaaaccgcgc 381  
Db 5946 CGACCTGGCGGAGCTGATCGCGGGGAGC 5919

Search completed: July 18, 2002, 11:22:05  
Job time: 13735 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:36:11 ; Search time 646.57 Seconds  
(without alignments)  
1080.756 Million cell updates/sec

Title: US-09-749-185-4

Perfect score: 407

Sequence: 1 atgagttctctgtctccga.....gaggacagagcgccggtcg 407

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	407	100.0	407	AAZ49729	Streptomyces albus
2	243.8	59.9	407	AAZ49730	Streptomyces golde
3	239	58.7	407	AAZ49728	Streptomyces grise
4	239	58.7	438	AAZ49727	Streptomyces grise
5	224.6	55.2	407	AAZ49731	Streptomyces netro
6	73	17.9	45624	AAZ49731	S. spinosa DNA fra
7	73	17.9	50000	AAZ49731	S. spinosa DNA fra
8	60.8	14.9	15872	AAZ49731	Streptomyces venez
9	60.8	14.9	15872	AAZ49731	S. venezuelae vep

10	59	14.5	1155	21	AAA47167	DNA encoding a Cal
11	58.4	14.3	18660	21	AAA58472	Nucleotide sequenc
12	56.2	13.8	3025	15	AAQ79402	Human NMDAR2 recep
13	56.2	13.8	3698	15	AAQ79400	Human NMDAR2 recep
14	56.2	13.8	3698	20	AAV82909	Human N-methyl-D-a
15	56.2	13.8	3698	21	AAA95032	Human N-methyl-D-a
16	56.2	13.8	3698	21	AAZ38723	Human clone NMDA22
17	56.2	13.8	3698	24	ABI99189	Human NMDA recepto
18	56.2	13.8	4002	20	AAV82915	Human N-methyl-D-a
19	56.2	13.8	4002	21	AAA95038	Human N-methyl-D-a
20	56.2	13.8	4002	21	AAZ38729	Human NMDAR2 subun
21	56.2	13.8	4002	24	ABI99194	Human NMDA2C enco
22	56.2	13.8	4017	20	AAV82913	Human N-methyl-D-a
23	56.2	13.8	4017	21	AAA95036	Human N-methyl-D-a
24	56.2	13.8	4017	21	AAZ38727	Human NMDAR2 subun
25	56.2	13.8	4017	24	ABI99192	Human NMDA2C enco
26	56.2	13.8	4053	20	AAV82912	Human N-methyl-D-a
27	56.2	13.8	4053	21	AAA95035	Human N-methyl-D-a
28	56.2	13.8	4053	21	AAZ38726	Human NMDAR2 subun
29	56.2	13.8	4053	24	ABI99191	Human NMDA2C enco
30	56.2	13.8	4068	20	AAV82889	Human N-methyl-D-a
31	56.2	13.8	4068	21	AAA95012	Human N-methyl-D-a
32	56.2	13.8	4068	21	AAZ38703	Human NMDAR2C subu
33	56.2	13.8	4068	24	ABI99169	Human NMDA recepto
34	56.2	13.8	4077	20	AAV82914	Human N-methyl-D-a
35	56.2	13.8	4077	21	AAA95037	Human N-methyl-D-a
36	56.2	13.8	4077	21	AAZ38728	Human NMDAR2 subun
37	56.2	13.8	4077	24	ABI99193	Human NMDA2C enco
38	56.2	13.8	4092	20	AAV82911	Human N-methyl-D-a
39	56.2	13.8	4092	21	AAA95034	Human N-methyl-D-a
40	56.2	13.8	4092	21	AAZ38725	Human NMDAR2 subun
41	56.2	13.8	4092	24	ABI99190	Human NMDA2C enco
42	56.2	13.8	4274	15	AAQ79407	Human NMDAR2 recep
43	56.2	13.8	4289	15	AAQ79405	Human NMDAR2 recep
44	56.2	13.8	4325	15	AAQ79404	Human NMDAR2 recep
45	56.2	13.8	4340	15	AAQ79372	Human N-methyl-D-a

#### ALIGNMENTS

RESULT 1  
AAZ49729  
ID AAZ49729 standard; DNA; 407 BP.  
XX  
AC AAZ49729;  
XX  
DT 18-APR-2000 (first entry)  
XX Streptomyces albus G ssgA gene.  
XX  
KW ssgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 3004;  
KW bioinsecticide; receptor agonist; antagonist; biomass; ds.  
XX  
OS Streptomyces albus G.  
XX  
FH Key Location/Qualifiers  
FT mat\_peptide 1..405  
FT /tag= a  
FT /label= SsgA\_protein  
XX  
PN WO200000613-A1.  
XX  
PD 06-JAN-2000.  
XX  
PF 25-JUN-1999; 99WO-NL00395.  
XX  
PR 26-JUN-1998; 98EP-0202148.  
XX  
PA (UYLE-) RIJKSUNIV LEIDEN.

```
PA (NEW-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.
XX
PI Van Wezel GP, Kraal B, Luiten RGM;
XX
DR WPI: 2000-147269/13.
DR P-PSDB; AAY44650.
XX
XX Reducing branching and enhancing fragmentation in filamentous
PT microorganisms used to improve their liquid culturing properties -
XX
XX Disclosure; Fig 5; 60pp; English.
XX
XX The present sequence is a DNA (ATCC 3004) encoding S. albus G SsgA
CC protein. SsgA reduces branching and fragment septation and enhances
CC fragmentation of mycelium in liquid culture resulting in lower viscosity
CC of culture broths. Filamentous bacteria can be transformed with ssgA
CC gene-containing plasmid to enhance the production of secondary
CC metabolites such as, antibiotics, antitumour agents, immunosuppressive
CC agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine
CC agents, herbicides, antiparasitic agents, ruminant growth promoters,
CC bioinsecticides, receptor agonists and antagonists and biomass.
XX
SQ Sequence 407 BP; 57 A; 150 C; 139 G; 61 T; 0 other;

Query Match 100.0%; Score 407; DB 21; Length 407;
Best Local Similarity 100.0%; Pred. No. 2.8e-61;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgagcttcctcgtccgaggagctcgcttcgcgatacccggtggagctgcggtacgag 60
DB 1 atgagcttcctcgtccgaggagctcgcttcgcgatacccggtggagctgcggtacgag 60
QY 61 accgtgatcgtacgcgtgcggtacgttcacatccaccccgagagcccggtacac 120
DB 61 accgtgatcgtacgcgtgcggtacgttcacatccaccccgagagcccggtacac 120
QY 121 tgggtcttcggcgtaacgtcgtggtcagggagcttcgagccgcggcgagcgcgac 180
DB 121 tgggtcttcggcgtaacgtcgtggtcagggagcttcgagccgcggcgagcgcgac 180
QY 181 gtccgggtctgcgcgtggtcgcgagcgagcgccacacgagggagtgacatcacctccaggtc 240
DB 181 gtccgggtctgcgcgtggtcgcgagcgagcgccacacgagggagtgacatcacctccaggtc 240
QY 241 ggctccgagcagcgctctcccgctcgcgaagcgccgctcgtcgccttcctcgaccgc 300
DB 241 ggctccgagcagcgctctcccgctcgcgaagcgccgctcgtcgccttcctcgaccgc 300
QY 361 gacgacgctctgacacgacccctcgcgagggagcagagcgccggctg 407
DB 361 gacgacgctctgacacgacccctcgcgagggagcagagcgccggctg 407

RESULT 2
AAZ49730
ID AAZ49730 standard; DNA; 407 BP.
XX
AC AAZ49730;
XX
XX 18-APR-2000 (first entry)
XX
XX Streptomyces goldeniensis ssgA gene.
XX
KW ssgA; liquid culture; filamentous bacteria; secondary metabolite;
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;
KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 21386;
KW bioinsecticide; receptor agonist; antagonist; biomass; ds.
```

```
XX Streptomyces goldeniensis.
XX
XX Key Location/Qualifiers
FT mat_peptide 1..405
FT /*tag= a
FT /label= SsgA_protein
XX
XX WO200000613-A1.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-NL00395.
XX
XX 26-JUN-1998; 98EP-0202148.
XX
XX (UYLE-) RIJKSUNIV LEIDEN.
XX (NEW-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.
XX
XX Van Wezel GP, Kraal B, Luiten RGM;
XX
XX WPI: 2000-147269/13.
XX P-PSDB; AAY44651.
XX
XX Reducing branching and enhancing fragmentation in filamentous
PT microorganisms used to improve their liquid culturing properties -
XX
XX Disclosure; Fig 5; 60pp; English.
XX
XX The present sequence is a DNA (ATCC 21386) encoding S. goldeniensis SsgA
CC protein. SsgA reduces branching and fragment septation and enhances
CC fragmentation of mycelium in liquid culture resulting in lower viscosity
CC of culture broths. Filamentous bacteria can be transformed with ssgA
CC gene-containing plasmid to enhance the production of secondary
CC metabolites such as, antibiotics, antitumour agents, immunosuppressive
CC agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine
CC agents, herbicides, antiparasitic agents, ruminant growth promoters,
CC bioinsecticides, receptor agonists and antagonists and biomass.
XX
XX Sequence 407 BP; 56 A; 135 C; 145 G; 71 T; 0 other;

Query Match 59.9%; Score 243.8; DB 21; Length 407;
Best Local Similarity 74.9%; Pred. No. 1.8e-33;
Matches 305; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 1 atgagcttcctcgtccgaggagctcgcttcgcgatacccggtggagctgcggtacgag 60
DB 1 atgagcttcctcgtccgaggagctcgcttcgcgatacccggtggagctgcggtacgag 60
QY 61 accgtgatcgtacgcgtgcggtcgcgagcttcacatccaccccgagagcccggtacac 120
DB 61 accgtgatcgtacgcgtgcggtcgcgagcttcacatccaccccgagagcccggtacac 120
QY 121 tgggtcttcggcgtaacgtcgtggtcagggagcttcgagccgcggcgagcgcgac 180
DB 121 tgggtcttcggcgtaacgtcgtggtcagggagcttcgagccgcggcgagcgcgac 180
QY 181 gtccgggtctgcgcgtggtcgcgagcgagcgccacacgagggagtgacatcacctccaggtc 240
DB 181 gtccacatcgcccgaccccgagagcttcggtcgcgaggtcgtcgtcgcgcgaggtg 240
QY 241 ggctccgagcagcgctcttcgcgctcgcgcaagcgccgctcgtcgccttcctcgaccgc 300
DB 241 ggagcgaccagcgatgttcgcgggtcgcgacgcgccgctgtggccttcctcgaccgc 300
QY 301 accgaccagggcttgctcgtcggcagcgagcgggcagcgcccgacttcgacagccacctc 360
DB 301 accgaccagggcttgctcgtcggcagcgagcggttcctcgcgacttcgacgcccgtc 360
QY 361 gacgacgctctgacacgacccctcgcgagggagcagagcgccggctg 407
DB 361 gacgagcgcttgacgacatccctggcgagggagcagacgcccggctg 407
```



```
SQ Sequence 438 BP; 68 A; 145 C; 153 G; 72 T; 0 other;

Query Match      58.7%; Score 239; DB 21; Length 438;
Best Local Similarity 74.2%; Pred. No. 1.2e-32;
Matches 302; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 1 atgagctctctgctccgagagctgccttcgcgcatcccggtggagctgcgatacag 60
DB 31 atgagctctctgctccgagagctgccttcgcgcatcccggtggagctgcgatacag 90
QY 61 acgctgacatcgatcgaggggtgcgctgacgttcacatcccgagagacgcccgggtcacc 120
DB 91 gtcggcatcgatgcacatcccgatgacgttcacatcccgagcgatgccctgtgacc 150
QY 121 tgggtctctggcgctgaactgctgtagggagctcctgacgcgcgagcgacgac 180
DB 151 tggcgcttcggcgagctgctgtagggagcggtcgaacgcccagcgagcgagcgat 210
QY 181 gtccgggtgctgcgcgggtggggcagacgcccacagggaggtgcacatcacctccaggtc 240
DB 211 gtgcacatcgcccgacgagccgagggcctcgagatgtccacatccggctccaggtc 270
QY 241 ggtccgagacgagcgctcttcctgcgctgcgcaagcgccgctgctgccttcctgcagcgc 300
DB 271 ggcgcgacgctgctgcttcctgcgctgcgcaagcgccgctgctgccttcctgcagcgc 330
QY 301 accgacagggctgctgcctgcgacgagcgaggggacacacccgacttcgacagccactc 360
DB 331 acgacaagctgctgcgcctgcgacgagcgaggggacacacgctggtgacttcgacggcaactc 390
QY 361 gacgacgctctgaaccgacgctgcgagagcagagcagagcgccgagctg 407
DB 391 gaggacgactggccgcatcctcgcgagagcagagcagagcagcggctg 437

RESULT 5
AAZ49731
ID AAZ49731 standard; DNA; 407 BP.
XX
AC AAZ49731;
XX
DT 18-APR-2000 (first entry)
XX
DE Streptomyces netropsis ssrA gene.
XX
KW ssrA; liquid culture; filamentous bacteria; secondary metabolite;
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;
KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23940;
KW bioinsecticide; receptor agonist; antagonist; biomass; ds.
XX
OS Streptomyces netropsis.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..405
FT /*tag= a
FT /label= SsrA_protein
XX
PN WO200000613-A1.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-NL00395.
XX
XX 26-JUN-1998; 98EP-0202148.
XX
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.
XX
XX Van Wezel GP, Kraal B, Luiten RGM;
XX
XX WPI; 2000-147269/13.
XX
```

```
DR P-PSDB; AAY44652.
XX
PT Reducing branching and enhancing fragmentation in filamentous
PT microorganisms used to improve their liquid culturing properties -
XX
XX Disclosure; Fig 5; 60pp; English.
XX
XX The present sequence is a DNA (ATCC 23940) encoding S. netropsis SsrA
CC protein. SsrA reduces branching and fragment septation and enhances
CC fragmentation of mycelium in liquid culture resulting in lower viscosity
CC of culture broths. Filamentous bacteria can be transformed with ssrA
CC gene-containing plasmid to enhance the production of secondary
CC metabolites such as, antibiotics, antitumour agents, immunosuppressive
CC agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine
CC agents, herbicides, antiparasitic agents, ruminant growth promoters,
CC bioinsecticides, receptor agonists and antagonists and biomass.
XX
SQ Sequence 407 BP; 63 A; 151 C; 131 G; 62 T; 0 other;

Query Match      55.2%; Score 224.6; DB 21; Length 407;
Best Local Similarity 72.0%; Pred. No. 3.3e-30;
Matches 293; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 1 atgagctctctgctccgagagctgccttcgcgcatcccggtggagctgcgatacag 60
DB 1 atgagctctctgctccgagagctgccttcgcgcatcccggtggagctgcgatacag 60
QY 61 acgctgacatcgatcgaggggtgcgctgacgttcacatcccgagagacgcccgggtcacc 120
DB 61 accgggatacctacgcggtgcgagatgaccttcacatcccgagagcgcctgtgacc 120
QY 121 tgggtctctggcgtagaactgctgtagggagctcctgagcgcgggcgagcgagcgac 180
DB 121 tggcgcttcggcgagctgctgtagggagctcctgagcgggatcaacgcccagcgagcgac 180
QY 181 gtccgggtgctgcgcgggtggggcagacgcccacagggaggtgcacatcacctccaggtc 240
DB 181 gtccacatcgcccgacgacgcccagggcctgctgcgagctctccatccggtccaggtg 240
QY 241 ggtccgagcagcgctcttcgcgctgcgcaagcgccgctgctccttcctgcagcgc 300
DB 241 ggcgcgagcgcgctcttcgcgctgcgcaagcgccgctgctccttcctgcagcgc 300
QY 301 accgacagggcttgcctgcctgcgagcgagcgggcagacacgaggtgcacatcacctccaggtc 360
DB 301 accgacaagtgcggtgcgctgcgtagaacagactctggtgacttcgagagcagcctg 360
QY 361 gacgacgctctgaaccgacgctgcgtagggagcagagcgccggctg 407
DB 361 gaggccgcgctcggaagatcctgcgagagcagagcagacgagcggtg 407

RESULT 6
AAF88315/c
ID AAF88315 standard; DNA; 45624 BP.
XX
AC AAF88315;
XX
DT 28-AUG-2001 (first entry)
XX
XX S. spinosa DNA fragment SEQ ID 4.
XX
XX Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; ds.
XX
OS Saccharopolyspora spinosa.
XX
XX DE19957268-A1.
XX
XX 08-MAR-2001.
XX
```



DE Streptomyces venezuelae polyketide synthase vep ORF1.  
 XX polyketide synthase; polyhydroxyalkanoate monomer synthase;  
 KW polyhydroxybutyrate; biodegradable polymer; vep gene;  
 KW metabolic engineering; ss.  
 XX Streptomyces venezuelae.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 20..13912  
 FT /\*tag= a  
 FT CDS 14056..14136  
 FT /\*tag= b  
 FT CDS 14148..15827  
 FT /\*tag= c  
 XX  
 PN WO9722711-A1.  
 XX  
 XX 26-JUN-1997.  
 XX  
 XX 18-DEC-1996; 96WO-US20119.  
 XX  
 XX 19-DEC-1995; 95US-0008847.  
 XX  
 XX (MINU ) UNIV MINNESOTA.  
 XX  
 XX Sherman DH, Williams MD, Xue Y;  
 XX  
 XX WPI; 1997-341701/31.  
 XX  
 XX P-PSDB; AAW19629-30 AND AAW00918.  
 XX  
 XX Expression cassettes for production of polyhydroxyalkanoate(s) -  
 PT provide wide range of biodegradable polymers for medical or  
 PT industrial use  
 XX  
 XX Claim 54; Fig 23; 91pp; English.  
 XX  
 XX Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide  
 CC synthase (PKS) gene cluster encoding a polyene of 12 carbons (see  
 CC also AAW19629-30 and AAW00918). It contains 5 PKS modules, with a 5'  
 CC loading module and a 3' end domain. Each of the sequenced modules  
 CC includes a keto-ACP, an acyltransferase, a dehydratase, a keto-  
 CC reductase and an acyl carrier protein domain. The gene cluster was  
 CC cloned using a heterologous hybridisation strategy from a genomic  
 CC DNA library. A novel expression cassette encoding the first module  
 CC from the vep gene cluster and module 7 from the Streptomyces typ  
 CC gene cluster has polyhydroxyalkanoate (PHA) monomer synthase  
 CC activity and can be used for PHA prodn. in host (esp. insect) cells  
 CC for use as a biodegradable polymer.  
 XX  
 XX Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;  
 SQ  
 Query Match 14.98; Score 60.8; DB 18; Length 15872;  
 Best Local Similarity 47.4%; Pred. No. 0.017;  
 Matches 182; Conservative 0; Mismatches 202; Indels 0; Gaps 0;  
 Qy 6 ctctctcgtccgagagctcgcttccgcacatcccggtgagctgagctgagctgagcagcgt 65  
 Db 1375 cgtcaccggtcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1434  
 Qy 66 cgatccgtacggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 125  
 Db 1435 cgaagcgagccgagctcggtcggtcggtcggtcggtcggtcggtcggtcggtcggt 1494  
 Qy 126 ctctggggtgagctggtggtggtggtggtggtggtggtggtggtggtggtggtggt 185  
 Db 1495 ggtcttcacgacggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1554  
 Qy 186 ggtctgctgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 245  
 Db 1555 cggcgtctccggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1614

Qy 246 cgagcagcgctctccgcgtcgcaagcgccgctgctgctcctcctcgaccgcaccca 305  
 Db 1615 gcgctcgctcctgttcagcgccaggggtgcccacgttacggcgatggatggtt 1674  
 Qy 306 ccagggtgtgctcggtcgagcggtggtggtggtggtggtggtggtggtggtggt 365  
 Db 1675 gtacggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1734  
 Qy 366 cgctctgaacgcgagcgtcgccga 389  
 Db 1735 cctcctgagcgggccccctcgccga 1758  
 RESULT 9  
 AAZ87283  
 ID AAZ87283 standard; DNA; 15872 BP.  
 XX  
 XX AAZ87283;  
 XX  
 XX 05-JUN-2000 (first entry)  
 XX  
 XX S. venezuelae vep ORF 1, SEQ ID NO:1.  
 XX  
 KW Desamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,  
 KW chronic obstructive pulmonary disease; respiratory inflammation;  
 KW hypercholesterolaemia; crop protection agent; ds.  
 XX  
 OS Streptomyces venezuelae AFCC15439.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 20..13912  
 FT /\*tag= a  
 FT CDS /product= "vep ORF 1 amino acid sequence #1 (AAZ77177)"  
 FT 14056..14151  
 FT /\*tag= b  
 FT CDS /product= "vep ORF 1 amino acid sequence #3 (AAZ77199)"  
 FT 14167..15827  
 FT /\*tag= c  
 FT CDS /product= "vep ORF 1 amino acid sequence #2 (AAZ77178)"  
 XX  
 XX WO200000620-A2.  
 XX  
 XX 06-JAN-2000.  
 XX  
 XX 25-JUN-1999; 99WO-US14398.  
 XX  
 XX 26-JUN-1998; 98US-0105537.  
 XX  
 XX (MINU ) UNIV MINNESOTA.  
 XX  
 XX Sherman DH, Liu H, Xue Y, Zhao L;  
 XX  
 XX WPI; 2000-160679/14.  
 XX  
 XX P-PSDB; AAZ77177, AAZ77178, AAZ77199.  
 XX  
 XX Desamine and macrolide biosynthetic gene clusters, useful for, e.g.  
 PT synthesis of methymycin and pikromycin -  
 XX  
 XX Example 3; Figure 23; 438pp; English.  
 XX  
 CC The invention relates to an isolated and purified nucleic acid segment  
 CC comprising a desamine biosynthetic gene cluster, a fragment or its  
 CC biologically active variant, where the nucleic acid sequence is not  
 CC derived from the erc gene cluster of Saccharopolyspora erythraea or  
 CC Streptomyces antibioticus. The invention also relates to a macrolide  
 CC biosynthetic gene cluster, or fragments thereof. The macrolide  
 CC biosynthetic gene cluster encodes proteins which synthesize methymycin,  
 CC pikromycin, neomethymycin, narbomycin or a combination of these  
 CC compounds. Recombinant or augmented cells comprising the desamine  
 CC and/or macrolide biosynthetic gene clusters are useful for the  
 CC production of biologically active macrolides. The macrolide biosynthetic









```

Db 432 gaccgggagggcgcgcgacgagcgctgtggtgacgagctcgcgagcgccgtg 491
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 283 ctgccttcttcgacgacgacgagggcttgcgtcgcgagcgagcgagcgacacgccc 342
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 492 ttgtgacctactcgcgcgagggagggcgaggtgctcttcgagggcgagcgagggc 551
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 343 gacttcgacagccacctcgcgagcgctctgaaccgagcctcgcgaggg 391
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 552 ggtctgtggggcccgccacgtgtgctgtgcccacactggcgctgg 600
      || | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
AAV82909
ID AAV82909 standard; cDNA; 3698 BP.
XX
AC AAV82909;
XX
DT 23-FEB-1999 (first entry)
XX
DE Human N-methyl-D-aspartate receptor subunit clone NMDA22.
XX
KW Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW NMDA-activated cation-selective ion channel; glutamate receptor; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 3..3698
FT FT /*tag= a
FT FT /product= NMDA_receptor_subunit
FT FT /note= "no termination codon"
XX
XX US5849895-A.
XX
XX 15-DEC-1998.
XX
XX 20-APR-1994; 94US-0231193.
XX
XX 20-APR-1994; 94US-0231193.
XX 20-APR-1993; 93US-0052449.
XX
XX (SIBI-) SIBIA NEUROSCIENCES INC.
XX
XX Daggett LP, Lu C;
XX
XX WPI; 1999-069812/06.
XX P-PSDB; AAW87503.
XX
XX DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
XX assembly of functional glutamate receptor subunits
XX
XX Example 3; Columns 253-262; 203pp; English.
XX
CC The present sequence encodes a human N-methyl-D-aspartate (NMDA)
CC receptor subunit (NMDAR). The nucleic acid sequence does not contain
CC the 366 5'-most nucleotides, has an additional 11 nucleotides between
CC nucleotides 1300 and 1301, nor the 15 nucleotides at positions
CC 1960-1974, nor the 1061 3' nucleotides, as set forth in AAV82889. The
CC cDNA sequence is derived from clone NMDA22. The NMDAR subunits contribute
CC to the formation of NMDA-activated cation-selective ion channels. In
CC addition to being useful for the production of NMDA receptor subunit
CC proteins, the nucleic acids are also useful as probes to identify and
CC isolate nucleic acids encoding related receptor subunits. Functional
CC glutamate receptors can be assembled from several NMDA receptor subunit
CC proteins of one type (homomeric) or from combinations of subunit proteins
CC of different types (heteromeric). The present invention also comprises
CC methods for using such receptor subunits to identify and characterise
CC compounds which affect the function of such receptors, e.g. agonists,
CC antagonists and modulators of glutamate receptor function. The invention
CC also comprises methods for determining whether unknown protein(s) are
CC functional as NMDA receptor subunits.
XX
```

```

SQ Sequence 3698 BP; 612 A; 1283 C; 1159 G; 644 T; 0 other;

Query Match 13.8%; Score 56.2; DB 20; Length 3698;
Best Local Similarity 47.8%; Pred. NO. 0.12;
Matches 166; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 43 gtgagctgcggtacagagacgctcgcgtacgaggtgcggtcgcgttcacaccccc 102
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 252 gtgtccctggagcagcagctgcaggtgcttcaagtgctggaagagctacgactggagc 311
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 103 ggagacgcccccggtgcacctgggtcttcggcggtgaactggtgcgagggagtgagtc 162
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 312 gcttgcgctcatcacacgctgcacccgggcacgcgtcttctcctggggcgctgcgc 371
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 163 gcgcggggcgacgagcgagctcgcgggtgcgcgggtggggcgagagcgccaccagggagt 222
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 372 gccgtgcgcgacgcccagccagctgagttggcgctgctgagctggcagctggaactg 431
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 223 cacatcacctccaggtcgcgctcgcgagcagcgctcttcgcgctcggcaaggcgccgctg 282
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 432 gacccgggagggcgcgcgcgacgacgagcgcgtcgtcgcagctcgcagcgcccgctg 491
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 283 ctgccttcttcgacgacgacgacgagggcttgcgtcgcgagcgagcgagcgacacgccc 342
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 492 ttgtggcctactcgcgcgagggagggcgaggtgctcttcgagggcgagcgagcgagggc 551
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 343 gacttcgacagccacctcgcgagcagcgtctgaaccgagcctcgcgaggg 391
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 552 ggtctgtggggcccgccacgtgtggtggtgcccacactggcgctgg 600
      || | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
AAV95032
ID AAV95032 standard; cDNA; 3698 BP.
XX
AC AAV95032;
XX
DT 23-FEB-2001 (first entry)
XX
DE Human N-methyl-D-aspartate receptor subunit NMDAR1A coding sequence #17.
XX
KW Human; N-methyl-D-aspartate receptor; NMDA; NMDAR1A; ionotropic;
KW glutamate receptor; drug screening; animal model; disease diagnosis;
KW genetic screening; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 3..3698
FT FT /*tag= a
FT FT /product= "NMDAR1A"
FT FT /transl_except= (pos:830..832,aa:Xaa)
FT FT /transl_except= (pos:912..914,aa:Xaa)
FT FT /transl_except= (pos:978..986,aa:Xaa)
FT FT /transl_except= (pos:1041..1043,aa:Xaa)
FT FT /transl_except= (pos:1266..1268,aa:Xaa)
FT FT /transl_except= (pos:1314..1316,aa:Xaa)
FT FT /transl_except= (pos:1392..1394,aa:Xaa)
FT FT /transl_except= (pos:1431..1433,aa:Xaa)
FT FT /transl_except= (pos:1503..1508,aa:Xaa)
FT FT /transl_except= (pos:1770..1772,aa:Xaa)
FT FT /transl_except= (pos:2136..2138,aa:Xaa)
FT FT /transl_except= (pos:2232..2237,aa:Xaa)
FT FT /transl_except= (pos:2529..2531,aa:Xaa)
FT FT /transl_except= (pos:2544..2546,aa:Xaa)
FT FT /transl_except= (pos:3027..3029,aa:Xaa)
FT FT /transl_except= (pos:3456..3458,aa:Xaa)
FT FT /transl_except= (pos:3627..3629,aa:Xaa)
FT FT /transl_except= (pos:3681..3683,aa:Xaa)
FT FT /note= "this sequence contains a number of in-frame
FT FT stop codons, represented by xaa in the protein sequence"
FT FT /partial
```

```
XX US61111091-A.
PN
XX
XX 29-AUG-2000.
PD
XX
XX 29-SEP-1997; 97US-0940086.
PF
XX
XX 20-APR-1994; 94US-0231193.
PR
XX 20-APR-1993; 93US-0052449.
PR
XX
XX (MERI ) MERCK & CO INC.
PA
XX
XX Daggett LP, Lu C;
PI
XX
XX WPI: 2000-578607/54.
DR
XX P-PSDB; AAB26239.
DR
XX
XX Novel DNA fragment encoding human N-methyl-D-aspartate receptor subunit
PT for identifying mutations and for developing drugs against various
PT disease states
PT
XX
XX Example 3; column 255-264; 205pp; English.
PS
XX
XX The present sequence is the coding sequence for a subunit (designated
CC NMDARIA) of the human N-methyl-D-aspartate (NMDA) receptor. These are
CC ionotropic glutamate receptors which contain cation-specific ligand-gated
CC ion channels. The sequence and its protein can be used in disease
CC diagnosis and in research to identify other, similar proteins. They can
CC also be used as probes, for example in genetic screening, and in drug
CC screening, as well as enabling the production of animal disease models.
XX
XX Sequence 3698 BP; 612 A; 1283 C; 1159 G; 644 T; 0 other;
SQ
```

Query Match 13.8%; Score 56.2; DB 21; Length 3698;  
Best Local Similarity 47.6%; Pred. No. 0.12;  
Matches 166; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

```
Qy 43 gtgagctgcggtacgagaccgtgatccgtacggtgctgagcttccacctccc 102
Db ||| ||| | | | | | | | | | | | | | | | | | | |
Qy 252 gtgtcccttgagcgagcagctcaggtgctgttcaaggtgctggaagatcacgactggagc 311
Db ||| ||| ||| ||| | | | | | | | | | | | | | | |
Qy 103 ggagacgcccgggtcacctgggtcttcggcgctgaactggtggtcggaggagtcctggac 162
Db | ||| ||| ||| | | | | | | | | | | | | | | |
Qy 312 gccctgcgcgtcatcaccagctgcacccggccacgcgctcttctctgaggcggtgcgc 371
Db | ||| ||| ||| | | | | | | | | | | | | | | |
Qy 163 gccgcggggcgagcgacgtccgggtgtgcccggtggggcgagcggccaccaggaggtg 222
Db ||| ||| ||| | | | | | | | | | | | | | | |
Qy 372 gccgtgcgcgacgcagccagctgagttggcggtgctgctggacgtggtcacgctggaactg 431
Db ||| ||| ||| | | | | | | | | | | | | | | |
Qy 223 cacatcacctccaggtggtcgcgagcagcgctcttccgctcggcaaggcgcgctg 282
Db || | | | | | | | | | | | | | | | | | | | |
Qy 432 gaccgggaggggccgcgcgcacgcagcgctgctgcccagctcgacgcgccctg 491
Db || | | | | | | | | | | | | | | | | | | | |
Qy 283 ctgccttctctcgaccgcaccgaccaggtgttgcgttcgctgcagcgagcgggcacacgcc 342
Db | | | | | | | | | | | | | | | | | | | | |
Qy 492 ttgtggcctactgctgcgcgagagggcaggtgtcttctgcgcgagggcgcgagccc 551
Db | | | | | | | | | | | | | | | | | | | | |
Qy 343 gacttcgacagccacctcgacgagcgtctgaaccgacgacctgcgcgagg 391
Db | | | | | | | | | | | | | | | | | | | | |
Qy 552 ggtctgtggggcccgccacgtgtggtgtgccccacctggcgctgg 600
Db | | | | | | | | | | | | | | | | | | | | |
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:24:51 ; Search time 165.56 seconds  
(without alignments)  
603.847 Million cell updates/sec

Title: US-09-749-185-4  
Perfect score: 407  
Sequence: 1 atgagctctctctccga.....gaggagcagcgccggtg 407

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	60.8	14.9	15872	4	US-09-105-537-1
2	56.6	13.9	474	2	US-08-403-852D-14
3	56.6	13.9	474	3	US-08-510-646B-14
4	56.6	13.9	474	4	US-09-231-818-14
5	56.2	13.8	3698	2	US-08-231-193A-43
6	56.2	13.8	3698	2	US-08-486-273A-43
7	56.2	13.8	3698	3	US-08-480-474-43
8	56.2	13.8	3698	3	US-08-940-086A-43
9	56.2	13.8	3698	4	US-08-940-035A-43
10	56.2	13.8	4002	2	US-08-231-193A-53
11	56.2	13.8	4002	2	US-08-486-273A-53
12	56.2	13.8	4002	3	US-08-480-474-53
13	56.2	13.8	4002	3	US-08-940-086A-53
14	56.2	13.8	4002	4	US-08-940-035A-53
15	56.2	13.8	4017	2	US-08-231-193A-49
16	56.2	13.8	4017	2	US-08-486-273A-49
17	56.2	13.8	4017	3	US-08-480-474-49
18	56.2	13.8	4017	3	US-08-940-086A-49
19	56.2	13.8	4017	4	US-08-940-035A-49
20	56.2	13.8	4053	2	US-08-231-193A-47
21	56.2	13.8	4053	2	US-08-486-273A-47
22	56.2	13.8	4053	3	US-08-480-474-47
23	56.2	13.8	4053	3	US-08-940-086A-47
24	56.2	13.8	4053	4	US-08-940-035A-47
25	56.2	13.8	4068	2	US-08-231-193A-5
26	56.2	13.8	4068	2	US-08-486-273A-5
27	56.2	13.8	4068	3	US-08-480-474-5

28	56.2	13.8	4068	3	US-08-940-086A-5	Sequence 5, Appli
29	56.2	13.8	4068	4	US-08-940-035A-5	Sequence 5, Appli
30	56.2	13.8	4077	2	US-08-231-193A-51	Sequence 51, Appl
31	56.2	13.8	4077	2	US-08-486-273A-51	Sequence 51, Appl
32	56.2	13.8	4077	3	US-08-480-474-51	Sequence 51, Appl
33	56.2	13.8	4077	3	US-08-940-086A-51	Sequence 51, Appl
34	56.2	13.8	4077	4	US-08-940-035A-51	Sequence 51, Appl
35	56.2	13.8	4092	2	US-08-231-193A-45	Sequence 45, Appl
36	56.2	13.8	4092	2	US-08-486-273A-45	Sequence 45, Appl
37	56.2	13.8	4092	3	US-08-480-474-45	Sequence 45, Appl
38	56.2	13.8	4092	3	US-08-940-086A-45	Sequence 45, Appl
39	56.2	13.8	4092	4	US-08-940-035A-45	Sequence 45, Appl
40	55.2	13.6	30001	1	US-08-125-468-1	Sequence 1, Appli
41	55.2	13.6	30001	2	US-08-474-933-1	Sequence 1, Appli
42	54.6	13.4	2219	3	US-08-510-646B-17	Sequence 17, Appl
43	52.8	13.0	441529	4	US-09-103-840A-1	Sequence 1, Appli
44	52.4	12.9	1931	2	US-09-130-114-2	Sequence 2, Appli
45	51	12.5	44377	2	US-08-804-227C-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-09-105-537-1  
; Sequence 1, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-1

Query Match	14.9%;	Score 60.8;	DB 4;	Length 15872;
Best Local Similarity	47.4%;	Pred. No. 0.00059;		
Matches 182;	Conservative	0;	Mismatches 202;	Indels 0; Gaps 0;
Qy	6	cttctctgctctccgagagctcgcttccgcatcccggtggagctgcgtacgagaccgt	65	
Db	1375	cgctaccggccggcggcgagggccctcgccgagccgcccgcgctcgcagagggccgt	1434	
Qy	66	cgatcgctacgcgggtgcgctgacgttccacctcccgagagacgcccgggtcacctgggt	125	
Db	1435	cgaaagcgaccggagctcgccgcgcactcgcccgctcgctggtcaccaccgtac	1494	
Qy	126	cttcggggcgtagactgctggtcgagggagctcctcgacgcgcggcgagcgagctccg	185	
Db	1495	ggtcttcacgaccggtcggtcgctcctcgcccgagccgcccgcgctcctcgagggct	1554	
Qy	186	gggtgctccgggtggggcagacggccaccagggaggtgcacatcacctccaggtcggtc	245	
Db	1555	cgccgcccctcgccgggagccgcccgcgcccgggtgtccacggcaccgcccccgg	1614	
Qy	246	cgagcaggcgctcttcgctcgccgcaaggcgccgctgctcgtctcctcgacgcgacga	305	
Db	1615	gcgctcgccgctcttcgagggccaggggtgtcccaacgtacggcgtggtgaggtt	1674	
Qy	306	ccagggttgcgtcgccgagcgaggggcacacccgacttcgacagccactcgacga	365	
Db	1675	gtacgcgcccaccgccccttcgacggccttcgacgcgtcgccgagcgaactggacc	1734	



```
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.pristinaespiralis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..474
; OTHER INFORMATION: /product= "Partie du gene Snbd"
;
US-08-510-646B-14

Query Match 13.9%; Score 56.6; DB 3; Length 474;
Best Local Similarity 48.9%; Pred. No. 0.0039;
Matches 152; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

Qy 86 tgacgttccacctcccgagagaccccggtcacctgggtcttcggtgagcgtgactgctgg 145
Db 161 TGGTGCTCCAGGCGCGGTGCGCGCCCTGCTCACCCGCGACGCGCGCGGACCATCC 220

Qy 146 tcgaggagtgctgagacgagcgcgagcgagcgagcgagcgagcgagcgagcgagcgag 205
Db 221 CCGTCGCGAGCGCGGTGCGCGCCCTGCTCACCCGCGACGCGCGCGGACCATCC 280

Qy 206 cggccaccagggagtgacatcacctccaggtcggtcgagcagcgagcgagcgagcgag 265
Db 281 TCGTCAACACCTGCTGCGCACCGACACCTCCGCGGACCCACCTTCGCGGAATCG 340

Qy 266 tcggcaaggcgccgtgctgcttccttccttccttccttccttccttccttccttccttc 325
Db 341 TCGCACCGGTGCGGAGGTGCGGAGGTGCGGAGGTGCGGAGGTGCGGAGGTGCGGAGGTG 400

Qy 326 gcgagcggggacacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 385
Db 401 AGCTCGTGAAGAGGTCAACCCGAGCGCTCCCTGGCCGCAACCGCGTCTTCCAGGTG 460

Qy 386 ccgaggagcag 396
Db 461 TCCTGGCGCTG 471

RESULT 4
; Sequence 14, Application US/09231818
; Patent No. 6171846
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanc, Francis
; APPLICANT: Crozet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Grey-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
```

```
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/231,818
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03806.0054-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.pristinaespiralis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..474
; OTHER INFORMATION: /product= "Partie du gene Snbd"
;
US-09-231-818-14

Query Match 13.9%; Score 56.6; DB 4; Length 474;
Best Local Similarity 48.9%; Pred. No. 0.0039;
Matches 152; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

Qy 86 tgacgttccacctcccgagagaccccggtcacctgggtggttcttcggtgagcgtgactgctgg 145
Db 161 TGGTGCTCCAGGCGCGGTGCGCGCCCTGCTCACCCGCGACGCGCGCGGACCATCC 220

Qy 146 tcgaggagtgctgagacgagcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 205
Db 221 CCGTCGCGAGCGCGGTGCGCGCCCTGCTCACCCGCGACGCGCGCGGACCATCC 280

Qy 206 cggccaccagggagtgacatcacctccaggtcggtcgagcagcgagcgagcgagcgagcgag 265
Db 281 TCGTCAACACCTGCTGCGCACCGACACCTCCGCGGACCCACCTTCGCGGAATCG 340

Qy 266 tcggcaaggcgccgtgctgcttccttccttccttccttccttccttccttccttccttccttc 325
Db 341 TCGCACCGGTGCGGAGGTGCGGAGGTGCGGAGGTGCGGAGGTGCGGAGGTGCGGAGGTG 400

Qy 326 gcgagcggggacacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 385
Db 401 AGCTCGTGAAGAGGTCAACCCGAGCGCTCCCTGGCCGCAACCGCGTCTTCCAGGTG 460

Qy 386 ccgaggagcag 396
Db 461 TCCTGGCGCTG 471

RESULT 5
US-08-231-193A-43
```

Oy	343:gacttcgacagccacctcgacgcctctgaaccggcagcctcgcgcagg 391
Dd	552:GGTCTGGTGCGGCCGACAGTCGTGGTGGTCCCAACTGGCGGTGG 600
 RESULT      6	
US-08-486-273A-43	
; Sequence 43, Application US/08486273A	
; Patent No. 5985586	
; GENERAL INFORMATION:	
; APPLICANT: Daggett, Lorrie P.	
; APPLICANT: Ellis, Steven B.	
; APPLICANT: Liaw, Chen W.	
; APPLICANT: Lu, Chin-Chun	
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA	
; TITLE OF INVENTION: SUBUNIT ENCODING SAME AND USES THEREFOR	
; NUMBER OF SEQUENCES: 63	
; CORRESPONDENCE ADDRESS:	
; ADDRESSEE: Brown, Martin, Haller & McClain	
; STREET: 1660 Union Street	
; CITY: San Diego	
; STATE: CA	
; COUNTRY: U.S.A.	
; ZIP: 92101-2926	
; COMPUTER READABLE FORM:	
; MEDIUM TYPE: Floppy disk	
; OPERATING SYSTEM: PC-DOS/MS-DOS	
; SOFTWARE: Patent In Release #1.0, Version #1.25	
; CURRENT APPLICATION DATA:	
; APPLICATION NUMBER: US/08/486,273A	
; FILING DATE: 06-JUN-1995	
; CLASSIFICATION: 435	
; PRIOR APPLICATION DATA:	
; APPLICATION NUMBER: US 08/231,193	
; FILING DATE: 20-APR-1994	
; CLASSIFICATION: 435	
; ATTORNEY/AGENT INFORMATION:	
; NAME: Seidman, Stephanie	
; REGISTRATION NUMBER: 33,779	
; REFERENCE/DOCKET NUMBER: 6362-9383B	
; TELECOMMUNICATION INFORMATION:	
; TELEPHONE: 619-238-0999	
; TELEFAX: 619-238-0062	
; INFORMATION FOR SEQ ID NO: 43:	
; SEQUENCE CHARACTERISTICS:	
; LENGTH: 3698 base pairs	
; TYPE: nucleic acid	
; STRANDEDNESS: single	
; TOPOLOGY: linear	
; MOLECULE TYPE: cDNA	
; FEATURE:	
; NAME/KEY: Coding Sequence	
; LOCATION: 3...3698	
US-08-231-193A-43	
Query Match            13.8%; Score 56.2; DB 2; Length 3698;	
Best Local Similarity   47.6%; Pred. No. 0.0044;	
Matches 166; Conservative 0; Mismatches 183; Indels 0; Gaps 0;	
Oy	43 gtgagtgctgggttacgagaccgtcatccgtatcggttgcggtgatcctcccc 102
Dd	252 GTGTCCCTGGAGCAGCAGTCAGTCTTTCAAGTGTCTGAAGTAGTACCAGTGGAGC 311
Oy	103 ggaagacccccggtcacctggtttcctgggtgtaactggttgtaggttgtcctggac 162
Dd	312 GCCTTCGCCTATCACACAGCCTCACCGGGCCAGCGCTCTTCTTCTGGAGGGGTGCGC 371
Oy	163 gccgggggcaagcgatcctgggtgtgccgggtgggggtgggggtgggggtgggggtg 222
Dd	372 GCCGTGCCCCGACGCCAGCATGAGTGGCGGCTGCTGAGAGTAGTACGACTGGAGC 431
Oy	223 cacataccctccaggctcggttcgagcagcgctcttcgagcgtcgcaaggcccagt 282
Dd	432 GACCCGGAGGGCGCGCGCCAGCAGCGCTCTGTCGCCAGTCTGAGCGCCCCTGTG 491
Oy	283 ctgcctctcagcgcacccagcaggggttctgcctcgagcagcagcgagcgccacagcc 342
Dd	492 TTGTGTCCTTACTCTGCGCGAGAGCGAGTGTCTTCTGCGAGGGCGCGCGAGCC 551









COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,474  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9382B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4002 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 189..3833  
US-08-480-474-53

Query Match 13.8%; Score 56.2; DB 3; Length 4002;  
Best Local Similarity 47.6%; Pred. No. 0.0044;  
Matches 166; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 43 gtgagctgcggtacgagaccgtcgatccgtacggtgacggtgacgtgacgttccacctcccc 102  
Db 618 GTGTCCTGGAGCAGCAGCTCAGGTGCTGTTCAAGGTGCTGGAAGAGTACGACTGGAGC 677  
Qy 103 ggagacgccccggttcacctggttcggtgacgtgacgtgacgtgacgtgacgtgacgtgac 162  
Db 678 GCCTTCGCCGTATCACCAGCTGCACCGCGCCAGCGCTCTTCTTGGAGGGGTGGCG 737  
Qy 163 gccgagggcgagcgagctcggttcggttcggttcggttcggttcggttcggttcggttcggttcg 222  
Db 738 GCCCTCGCGCAGCGCAGCAGCTGAGTTGGCGGCTGCTGGAGCTGGTTCACCTGGAAGT 797  
Qy 223 cacatcacctccaggtcgctccgagcagcgctcttccgctcgctcgctcgctcgctcgctcgct 282  
Db 798 GACCGGAGGGCGCGCGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857  
Qy 283 ctgccttcctcgaccgacccgacccgacccgacccgacccgacccgacccgacccgacccgaccc 342  
Db 858 TTTGTGGCTTACTGCTCGCGCAGGAGCGCGAGGTGCTTTCGCCGAGCGCGCGCAGGCC 917  
Qy 343 gacttcagaccacccctcgacgagcgtctgaacccgagcgtctcgccgaggtg 391  
Db 918 GGTCTGGTGGGGCGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 966

## RESULT 13

US-08-940-086A-53  
; Sequence 53, Application US/08940086A  
; Patent No. 6111091  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor

CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/940,086A  
FILING DATE: 29-SEPT-97  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/231,193  
FILING DATE: 20-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/052,449  
FILING DATE: 20-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 24735-9383C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 450-8400  
TELEFAX: (619) 450-8499  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4002 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 189..3833  
US-08-940-086A-53

Query Match 13.8%; Score 56.2; DB 3; Length 4002;  
Best Local Similarity 47.6%; Pred. No. 0.0044;  
Matches 166; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 43 gtgagctgcggtacgagaccgtcgatccgtacggtgacggtgacgtgacgttccacctcccc 102  
Db 618 GTGTCCTGGAGCAGCAGCTCAGGTGCTGTTCAAGGTGCTGGAAGAGTACGACTGGAGC 677  
Qy 103 ggagacgccccggttcacctggttcggtgacgtgacgtgacgtgacgtgacgtgacgtgac 162  
Db 678 GCCTTCGCCGTATCACCAGCTGCACCGCGCCAGCGCTCTTCTTGGAGGGGTGGCG 737  
Qy 163 gccgagggcgagcgagctcggttcggttcggttcggttcggttcggttcggttcggttcggttcg 222  
Db 738 GCCCTCGCGCAGCGCAGCAGCTGAGTTGGCGGCTGCTGGAGCTGGTTCACCTGGAAGT 797  
Qy 223 cacatcacctccaggtcgctccgagcagcgctcttccgctcgctcgctcgctcgctcgctcgct 282  
Db 798 GACCGGAGGGCGCGCGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857  
Qy 283 ctgccttcctcgaccgacccgacccgacccgacccgacccgacccgacccgacccgacccgaccc 342  
Db 858 TTTGTGGCTTACTGCTCGCGCAGGAGCGCGAGGTGCTTTCGCCGAGCGCGCGCAGGCC 917  
Qy 343 gacttcagaccacccctcgacgagcgtctgaacccgagcgtctcgccgaggtg 391  
Db 918 GGTCTGGTGGGGCGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 966

## RESULT 14

US-08-940-035A-53  
; Sequence 53, Application US/08940035A  
; Patent No. 6316611  
; GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Liaw, Chen W.  
APPLICANT: Lu, Chin-Chun  
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/940,035A  
FILING DATE: 29-SEPT-97  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/231,193  
FILING DATE: 20-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/052,449  
FILING DATE: 20-APR-1993  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9383E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4002 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 189..3833  
US-08-940-035A-53

Query Match 13.8%; Score 56.2; DB 4; Length 4002;  
Best Local Similarity 47.6%; Pred. No. 0.0044;  
Matches 166; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 43 gtggagctgcgtacgagaccgtcgatccgtacgcggtgcggtgacgttcacacccccc 102  
Db 618 GTGTCCTCGGAGCAGCAGCTCAGGTGCTGTTCAAGGTGCTGGAAGACTACGACTGGAGC 677  
Qy 103 ggagacgccccggttcacctggtcttcggtggaactgctggtcgagggagtcctggac 162  
Db 678 GCCTTCGCGCTCATCACCAGCCTGCACCGGGCCACGCGCTCTTCTCGAGGGCGTGGCG 737  
Qy 163 gcccgggcgagcggcagctgcgggtctgccggtggggcagcagccaccagggagtg 222  
Db 738 GCGGTCGCGGAGCCGAGCAGCTGAGTTGGCGGCTGCTGGAGCTGGTCACTGGAACCTG 797  
Qy 223 cacatcacctccagtgctccgagcagggctcttcccgctcggaaggcgccgctg 282  
Db 798 GACCGGGAGGGCGCGCGCAGCAGCGCTGCTGCGCCAGCTCGACGCGCGCTG 857  
Qy 283 ctgcgcttcctcgaccgcgacccagcagggcttgctgctcgagcagcgggacacgcc 342  
Db 858 TTTGTGGCTACTGTCTCGCGGAGGAGCCGAGGTGCTCTTTCGCGGAGGCGCGAGGCC 917

Qy 343 gacttcgacagccacccctcgacgagctctgaaccgagcagcctcgccgag 391  
Db 918 GGTCTGTTGGGCGCGGCACGTGTGGCTGTGCTGCCCAACCTGGCGCTGS 966  
RESULT 15  
US-08-231-193A-49  
Sequence 49, Application US/08231193A  
Patent No. 5849895  
GENERAL INFORMATION:  
APPLICANT: Daggett, Lorrie P.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Liaw, Chen W.  
APPLICANT: Lu, Chin-Chun  
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/231,193A  
FILING DATE: 20-APR-1994  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/052,459  
FILING DATE: 20-APR-1993  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9383  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4017 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 189..3848  
US-08-231-193A-49

Query Match 13.8%; Score 56.2; DB 2; Length 4017;  
Best Local Similarity 47.6%; Pred. No. 0.0044;  
Matches 166; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 43 gtggagctgcgtacgagaccgtcgatccgtacgcggtgcggtgacgttcacacccccc 102  
Db 618 GTGTCCTCGGAGCAGCAGCTCAGGTGCTGTTCAAGGTGCTGGAAGACTACGACTGGAGC 677  
Qy 103 ggagacgccccggttcacctggtcttcggtggaactgctggtcgagggagtcctggac 162  
Db 678 GCCTTCGCGCTCATCACCAGCCTGCACCGGGCCACGCGCTCTTCTCGAGGGCGTGGCG 737  
Qy 163 gcccgggcgagcggcagcagtcgggtctgccggtggggcagcagccaccagggagtg 222  
Db 738 GCGGTCGCGGAGCCGAGCAGCTGAGTTGGCGGCTGCTGGAGCTGGTCACTGGAACCTG 797

Qy 223 cacatcacccctccaggtcgggtccgagcagcgctcttccgctcggcaaggcgccgctg 282  
Db 798 GACCCGGGAGGGCGCGCGCACGACGCGCTGCTGCCAGCTCGACGCGCCCGTG 857  
Qy 283 ctgccttctctgaccgcaccgacccagggcttctcgtcggcagcgagcgggcacacgcc 342  
Db 858 TTGTGGCCTACTGCTCGCGGAGGAGCGGAGGTGCTTTCGCCGAGCGCGCGCAGGCC 917  
Qy 343 gacttcgacagccacacctcgacgctctgaaccgagcctcgccgagg 391  
Db 918 GGTCGTGGGGCCCCGCCACGCTGTGCTGTGCCCAACCTGGCGCTGG 966

Search completed: July 18, 2002, 11:25:01  
Job time: 10891 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 10:12:46 ; Search time 4821.96 Seconds  
(without alignments)  
1139.217 Million cell updates/sec

Title: US-09-749-185-4  
Perfect score: 407  
Sequence: 1 atgagcttcctgcgtctccga.....gaggagcagagcgccggctg 407

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
c 1	66	16.2	925	12	CNS0091P
2	64.2	15.8	925	12	CNS0091P
3	64.2	15.8	935	12	CNS006XX
c 4	58.8	14.4	390	12	AQ846145
5	58.8	14.4	554	10	BE496986
6	58.6	14.4	605	10	BE593775
7	58.4	14.3	566	9	AV938413
8	57.8	14.2	623	9	AL508175
c 9	57.4	14.1	844	12	CNS0052P
10	57.2	14.1	532	10	BE603206
11	57.2	14.1	761	10	BF266913
12	57.2	14.1	1152	12	AG076818
c 13	55.6	13.7	474	12	AQ851336
14	55.2	13.6	572	10	BM324683
15	55.2	13.6	615	10	BM325544
16	55	13.5	1516	10	BG809984
17	54.8	13.5	636	10	BE490147

c 18	54.8	13.5	935	12	CNS006XX
19	54.6	13.4	720	9	BE060622
c 20	54.4	13.4	556	12	AQ851215
21	54.2	13.3	496	10	EG263173
c 22	53.8	13.2	502	12	AQ848096
23	53.6	13.2	678	10	BI726984
c 24	53.4	13.1	1203	12	CNS015Y4
25	53.2	13.1	556	9	AW566001
c 26	53.2	13.1	707	10	BG366417
27	53.2	13.1	932	12	CNS0072Q
c 28	52.6	12.9	500	9	BE230578
29	52.6	12.9	570	9	AI833435
c 30	52.6	12.9	1161	12	AG049228
31	52.4	12.9	445	12	AQ849292
32	52.4	12.9	624	10	BI953854
33	52.4	12.9	663	9	AL505998
34	52.4	12.9	700	9	AL509103
35	52.2	12.8	585	10	BF255377
36	52	12.8	390	10	BF483973
37	52	12.8	473	10	BG357234
38	52	12.8	537	10	BE592130
39	52	12.8	605	10	EG357951
40	52	12.8	932	12	CNS0072Q
c 41	51.8	12.7	390	10	BI511102
42	51.8	12.7	412	10	BG263550
43	51.8	12.7	440	12	AQ852321
44	51.8	12.7	479	9	AU181382
45	51.6	12.7	428	10	BF484459

ALIGNMENTS

RESULT 1  
CNS0091P/c  
LOCUS  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL053013  
VERSION AL053013.1 GI:4934461  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mannosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
FEATURES  
source  
Location/Qualifiers  
1. .925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"

```
/clone="BACR19D16"
/notes="end : TET3"
BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match 16.2%; Score 66; DB 12; Length 925;
Best Local Similarity 11.8%; Pred. No. 0.48;
Matches 40; Conservative 176; Mismatches 124; Indels 0; Gaps 0;

Qy 66 cgatccgtagcggtgagctgacgttccacctcccccggagagacgcccgttaccctgggt 125
Db 925 CSBSCSCSCSBSCSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 866
Qy 126 cttcgggcgtagctgtgtgtagggagtgcttgagcgcgcggcgagcgagcagctcg 185
Db 865 SSGTSSACVKNASSCCCGCMABCCMSSSSCCGSASARGVKVRASGGAGKRGGS 806
Qy 186 ggttcgcgggtggggcagcgccacccagggaggtgacacatccacctcgagtcggtc 245
Db 805 GGAASHSSSSACBSSSSSCSACWSASSSSASSSRSRSGGGAGGSGAGSSSSSSSS 746
Qy 246 cgagcagcgtcttcgcgtcggaaggcgctgctgcgtcttccttcgacgcaccga 305
Db 745 ASAGVSVSSASSSSSSSSSSSVSSVSSMSSCSBSSSSASASSSSSSSSASCA SCC 686
Qy 306 ccagggttgcgtcgagcagcgccacacgcgcgacttcgacagccacactcgacga 365
Db 685 TSMSCSCTSNASMAARSSSSSSSSSSSSSMASASSSSASSSSSSSSSSSSSGSACGB 626
Qy 366 cgtctgaacgcagcgtccgagagcagagcgccgc 405
Db 625 SMSSGGGSGSVASGMSVSSVSSGGRSGSGGGGVGGS 586

RESULT 2
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
Location/Qualifiers
1..925
/organism="Drosophila melanogaster"

/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR19D16"
/notes="end : TET3"
BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match 15.8%; Score 64.2; DB 12; Length 925;
Best Local Similarity 13.4%; Pred. No. 0.88;
Matches 49; Conservative 177; Mismatches 141; Indels 0; Gaps 0;

Qy 34 cgcatacccggtgagtggtgtagcagacgctgcgtacgcggtcgcgctgagcgttc 93
Db 558 SGYKGCSCSGSGSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSC 617
Qy 94 cactccccggagacaccccggtcactgggtcttggcgctgaactgctggcgagggga 153
Db 618 SCCCSKSVCGTSCSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 677
Qy 154 gtctggagcgcgcggcgagcgacgtcgcgtgtctgcccggtggggcgagagcgccac 213
Db 678 SGSGSNWAGGSGGTGTSSSSSSSSSSTSSSVSSGSKSTSSGSSBSSSSSSSSSTS 737
Qy 214 agggaggtgcataccctccaggtcggtcgctcgagcagcgctcttcgcgtcggaag 273
Db 738 SBBSCSTSSSSSSSSSSSTSCCTCCCSYSSSSSTSSSSSTSWGSTSGSSSSSVTSS 797
Qy 274 gcgcctgctgcctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 333
Db 798 SDSTSTCCSCCCYMCCTCYBMBCTYSTSCGGSSSSSSSGKGGVTKCCGCGGSSSNGMB 857
Qy 334 gcacgcgcgacttcgacagccacttcgacgcgacttcgacgcgcgcgcgcgcgcgcg 393
Db 858 GTSSACSSSSSSSSSVSSSSSKSSASSSSSVSSSGSVSSSSASKSSSSGSGSG 917
Qy 394 cagagcg 400
Db 918 SGSGSVS 924

RESULT 3
CNS006XK 935 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL066051
VERSION AL066051.1 GI:4945019
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
Location/Qualifiers
1..925
/organism="Drosophila melanogaster"
```



and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source  
Location/Qualifiers  
1..935  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-928"  
/clone="BACRI4N09"  
/note="end : T7"

BASE COUNT 257 a 170 c 162 g 96 t 250 others  
ORIGIN

Query Match 15.8%; Score 64.2; DB 12; Length 935;  
Best Local Similarity 29.5%; Pred. No. 0.88;  
Matches 113; Conservative 104; Mismatches 165; Indels 1; Gaps 1;

QY 23 agctcgcttcgcacccccgggtgagctgcgtacgagaccgtcgatccgtacgcgtgc 82  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 491 ANNAVSTGTMCDDMMVCGCGSCMCRTSSGSGTGCTCCMYMSSSVSCSCSGTCCGTCSS 550  
QY 83 ggctgaattccacctcccccggagacccccgggtcacctcgggtcttctggcggtgaactgc 142  
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 551 CTSKMCSCYTGCKCGCGCGCTSCSSSSCCBBSYTCBTCTKCSGCGSGGCTGC 610  
QY 143 tggtcgaggagctctgacgc 202  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 611 CGGGGCG 670  
QY 203 agacggccacaggaggtgcacatcacctctccaggtcgcgtcgcgcgcgcgcgcgcgcgcgcgc 262  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 671 GGGCGGCG 729  
QY 263 gcctcgcaaggcgctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgt 322  
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 730 SSGSGCG 789  
QY 323 gcagcgagcgggcacgc 382  
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 790 CSGCG 849  
QY 383 tcgccgaggagcagcgccgc 405  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 850 GCGCGSGGSCCG 872

## RESULT 4

AQ846145/c  
LOCUS  
DEFINITION  
LMAJFV1\_lm15c01.y1 Leishmania major FV1 random genomic library  
Leishmania major genomic clone LMAJFV1\_lm15c01 5' similar to  
contains element 212bp.2 leishmania repetitive element ;, DNA  
sequence.

## ACCESSION

VERSION AQ846145

## KEYWORDS

GSS.

## SOURCE

Leishmania major.

## ORGANISM

Leishmania major  
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;

## REFERENCE

1 (bases 1 to 390)

## AUTHORS

Apopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L.,  
Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A., Li, L.,  
Blistain, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko,  
I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M.,  
Harvey, N., McCann, R., Tsagarishvili, R., Williams, T., Jackson, Y.,  
Bowers, Y., Swaller, T., Waterston, R., Wilson, R., and Beverley, S.M.  
A survey of the Leishmania major Friedlin strain V1 genome by  
shotgun sequencing: a resource for DNA microarrays and expression  
profiling

## TITLE

Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)

## JOURNAL

MEDLINE  
21192569

## COMMENT

Contact: Apopyants, NS / Beverley, SM  
WashU Leishmania Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Library construction: Natalia S. Apopyants, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
If using this information please cite:  
N.S. Apopyants and S.M. Beverley 'A survey of the Leishmania major  
Friedlin strain V1 genome by shotgun sequencing' and the Washington  
University Genome Sequencing Center For information on obtaining  
clone material please contact: Natalia S. Apopyants Ph.D.  
(natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.  
(beverley@borcim.wustl.edu)  
Seq primer: -40RP from Gbco  
Class: shotgun  
High quality sequence stop: 302.  
Location/Qualifiers  
source  
1..390  
/organism="Leishmania major"  
/strain="Friedlin strain V1"  
/db\_xref="taxon:5664"  
/clone="LMAJFV1\_lm15c01"  
/lab\_host="TOP10 (Invitrogen)"  
/note="Vector: pZero-2 (Invitrogen); Site:1: EcoRV;  
Genomic DNA was isolated from stationary phase cells. For  
this library, DNA was sheared to give a tight size  
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA  
polymerase, dephosphorylated with Shrimp Alkaline  
Phosphatase and ligated into pZero-2 vector's EcoRV site."  
BASE COUNT 71 a 122 c 167 g 30 t  
ORIGIN

Query Match 14.4%; Score 58.8; DB 12; Length 390;  
Best Local Similarity 48.8%; Pred. No. 5.6;  
Matches 159; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 63 cgtcgatccgtacgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgc 122  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 382 CCTCGTCTGCTGCGCGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 323  
QY 123 ggtcttcggcggtgaactcgtggtcgagggagtcgtgacccgcgcgcgcgcgcgcgcgcgcgcgc 182  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 322 CGCGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 263  
QY 183 ccgggtctgcgcgtgc 242  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 262 GCAGCTCGCGCGCGTGGCGTCTCGTCTGCGCGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 203  
QY 243 ctccgagcagcgctctcttcgc 302  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 202 CGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143  
QY 303 cgaccagggttgctcgtcgc 362  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 142 CGTGTGCGCGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 83  
QY 363 cgacgctcgaacgc 388  
Db 82 CCAGCAGCAGTCCGCGCGCGTCCGCGG 57

## RESULT 5

BE496986  
LOCUS  
DEFINITION  
WHE0763\_A05\_A09S Wheat heat-stressed seedling cDNA library  
Triticum aestivum cDNA clone WHE0763\_A05\_A09, mRNA sequence.  
ACCESSION  
BE496986  
VERSION  
BE496986.1 GI:9695603



Qy	342	cgactctgcagccacctgcagcagctcttgaaaccgcagccctcgcagagcagcgcgc	401
Db	311	GGTGAACGATCGTGGCGCGGACGCGCAACTCGCGTGGCCAAGGATATGGCGC	370
Qy	402	cggc	405
Db	371	GGGC	374
RESULT	8		
AL508175			
LOCUS	AL508175	Hordeum vulgare	
DEFINITION	AL508175	Hordeum vulgare cDNA clone HY07P24V 5', mRNA sequence.	
ACCESSION	AL508175		
VERSION	AL508175.1	GI:12034390	
KEYWORDS	EST.		
SOURCE	barley.		
ORGANISM	Hordeum vulgare		
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Hordeum.			
1 (bases 1 to 623)			
Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.			
EST sequencing and analysis in barley			
Unpublished (2000)			
Contact: Michalek W			
Institute for Plant Genetics and Crop Plant Research			
Corrensstr.3, D-06466 Gatersleben, Germany			
Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de			
Seq primer: T3 primer for 5'end.			
Location/Qualifiers			
1..623			
/organism="Hordeum vulgare"			
/cultivar="Barke"			
/db_xref="taxon:4513"			
/clone="HY07P24V"			
/clone_lib="Hordeum vulgare Barke developing caryopsis (3.-15.DAP)"			
/tissue_type="developing caryopsis (3.-15.DAP)"			
/lab_host="XL0LP"			
/note="Vector: plasmid pBK-CMW; Site_1: EcoRI; Site_2: XhoI; mRNA was made from developing caryopsis (3.-15.DAP) of spring barley variety 'Barke', a high quality malting variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI (3'-end of cDNA). NOTE: Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter. Average insert size is 1 kb. Sequence trimming: Vector sequences and sequence ends were trimmed from the 5'- and 3'-end until a 50 bp window contains less than two ambiguities. The maximum length was set to 700 bp."			
BASE COUNT	126 a	203 c	200 g
ORIGIN		91 t	3 others
Query Match	14.2%	Score 57.8; DB 9; Length 623;	
Best Local Similarity	46.9%	Prod. No. 7.8;	
Matches 179; Conservative	0; Mismatches 203; Indels	0; Gaps	0;
Qy	24	gctgccttcgcgacccggtggagctcggtacgacacgctcgatccgtacgcggtgcg	83
Db	59	GCTCTGCATGNGTCCGAGTTTCGACCTCGCCCTGGACCTCTGCACAGAGGGTCAACAA	118
Qy	84	gctgacgttcacactcccggagagacgcccgggtcacctgggtcttctggggtgaactgct	143
Db	119	GATGAAGATGATTTCGGCCCTCAACAGTCGGCGTTCGACATCGCGCGCACGCTCGCGCGCA	178
Qy	144	ggtcagggagctcctgacgacgcgcggcgacgcgcgagctcggtctgtgcccggtgggagca	203
Db	179	CGAGTACATTGGCAATGTCAGGCCGCGAGGTGCTCGACGACWTCTCTCGTACCCCGCGCA	238
Qy	204	gacggccaccaggaggtgtcacatcacctccaggtcggtccgagcagcagcgctcttcgcg	263



and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinjofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/gppages/bgn/31/cover.html>) "

Db	470	CGAAGCGGTTCGTTTCAGAGTGGACGCGATCGTGGGCGGACGCGGCGCAACTCGCGGT	529
Qy	384	cgcgaggagcagagccggc	405
Db	530	GGCCAAAGATATGGCGGGG	551
RESULT 12			
AG076818			
LOCUS	AG076818	1152 bp	DNA linear GSS 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-071C05.R, genomic survey sequence.		
ACCESSION	AG076818		
VERSION	AG076818.1 GI:16628620		
KEYWORDS	GSS; GSS (genome survey sequence).		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-071C05.R.		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.		
AUTHORS	1. (sites)		
TITLE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
JOURNAL	BAC end sequences of Library PTB		
REFERENCE	Unpublished		
AUTHORS	2. (bases 1 to 1152)		
TITLE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
JOURNAL	Direct Submission		
COMMENT	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpses@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
PRIMERS	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.		
Sequencing:	ML3Rev		
LIBRARY			
Vector	: pKS145		
R.Site 1	: SacI		
R.Site 2	: SacI.		
Location/Qualifiers			
1.	.1152		
/organism="Pan troglodytes"			
/db_xref="taxon:9598"			
/clone="PTB-071C05.R"			
/sex="male"			
/cell_type="lymphoblast"			
/clone_lib="PTB Chimpanzee Male BAC Library"			
BASE COUNT	34 a	488 c	514 g 23 t 93 others
ORIGIN			
Query Match	14.1%; Score 57.2; DB 12; Length 1152;		
Best Local Similarity	46.5%; Pred.No.9.3;		
Matches 152; Conservative	0;	Mismatches 175;	Indels 0; Gaps 0;
Qy	79	gtcgcggtgacgtttccacctcccgagagaccccggtcacctggttcttcgggcgtgaa	138
Db	502	CGCGAGGGGGNGGCGCCCGCGCGCGCCCGGNGGCGCGGCGGCGGCGGCGGCG	561
Qy	139	ctgtctgtctgaggagtccttgagcgcgcggcgacgcgcgtccgggtctgcgcggtg	198
Db	562	GGGGGGGGGGCGCCNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGG	621
Qy	199	gggcagagcgccaccaggaggtacacatcacctccaggtcgactccgacgagcagctc	258
Db	622	CGCGCGCGCGCGGGCGCNCGCGCCCGCGCGCGCCCGCGCGCCCGCGCGCGCGCG	681
Qy	259	tccgcgtctggcaaggcgccgctgctgccttcttcgacccagccagcaggggttgcg	318

/note="Vector: pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; Four-week-old sorghum seedlings were sprayed with spore suspension prepared from 3-week-old FRM421, a sorghum isolate of the anthracnose pathogen *Colletotrichum graminicola*. Inoculated plants were kept in a 25 C dark growth chamber with 100% relative humidity for 24 hr, followed by 12/12 hr of light/dark cycle at 25 C with 90% relative humidity for another 24 hr. All leaves were harvested and quick frozen with liquid nitrogen and stored in a -80 C freezer. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

```
BASE COUNT      93 a 182 c 218 g 79 t
ORIGIN

Query Match      13.6%; Score 55.2; DB 10; Length 572;
Best Local Similarity 49.3%; Pred. No. 19;
Matches 144; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy 114 ggtcaacctgggtcttcggggtgaaactgctggtcgagggagtccttgagcgcgcgggga 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 GATGCCCAAGGTGCGACCTCCGCGGGGTGCTGCCCGGGGCGCAGCGTGGGATGCCGCCG 117

Qy 174 cggcgacgtccgggtctcccggtgggacagcggccaccaggagggtgcacatcacct 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 CGCCCGGTGACCGCGTCCATGTTGGCGCACGGCTCGCTGCTGGCGCAGCAGCGCT 177

Qy 234 ccagtcggtccgagcagcgctcttcggtcgcaagcgccgctgctgcctctct 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 CGGCGGACCTCCGCGGGCGCTCTTCTCCGTCGCCCTGCCGAGCTCTTCGCGCTCC 237

Qy 294 cgaccgcaccgacccgggtctgctcgagcagcgagggggcacacgcgcgacttcgacag 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TCTCGAGGTCAAGCAGCGACGGTGTCTCGAACAAGGGCAGTTTCAGAGGCTACATCGCCA 297

Qy 354 ccactcgacagcgtctgaaccgagcgtccgagagagagagcgcgcgcgc 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 GCAGCCCGCATGAACCTGGGAGAGCTCCGCTCGGGGAGCCGCCACCGAGCC 349

RESULT 15
BM325544          615 bp      mRNA      linear      EST 04-JAN-2002
LOCUS            PIC1_45_H11_b1_A002 Pathogen-infected compatible 1 (PIC1) Sorghum
DEFINITION       bicolor cDNA, mRNA sequence.
ACCESSION        BM325544
VERSION           BM325544.1 GI:18064681
KEYWORDS          EST.
SOURCE            sorghum.
ORGANISM          Sorghum bicolor
                  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
                  clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE         1 (bases 1 to 615)
AUTHORS           Cordonnier-Pratt,M.-M., Gingle,A., Fang,G.C., Dean,R., Wing,R.,
                  Sudman,M. and Pratt,L.H.
TITLE             An EST database from Sorghum: plants infected with a compatible
                  pathogen
JOURNAL           Unpublished (2002)
COMMENT           Contact: Cordonnier-Pratt MM
                  Department of Botany
                  The University of Georgia
                  Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
                  Tel: 706 542 1860
                  Fax: 706 542 1805
                  Email: mmpratt@uga.edu
                  Sequences have been trimmed to exclude PolyA, vector, and regions
                  below Phred quality 16. The threshold for highest quality sequence
                  is 20. Three-prime sequences, which are obtained with PolyTmix or
                  T7 sequencing primer, are presented as the reverse complement.
                  Seq primer: JEN REV
                  High quality sequence stop: 469
POLYA-No.         Location/Qualifiers
1..572
/cultivar="B7x623"
/db_xref="taxon:4558"
/clone_lib="Pathogen-infected compatible 1 (PIC1)"
/tissue_type="Leaves"
/dev_stage="4-week-old seedlings infected with
Colletotrichum graminicola"
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:15:01 ; Search time 130.38 Seconds  
(without alignments)  
115.010 Million cell updates/sec

Title: US-09-749-185-5

Perfect score: 687

Sequence: 1 MSFLVSEELAFRIPELVRLYE.....FDSHLDALNRLAEQSQAG 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*

1: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSS/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	687	100.0	135	21	Streptomyces albus
2	481	70.0	135	21	Streptomyces goide
3	454	67.5	135	21	Streptomyces grise
4	446	64.9	135	21	Streptomyces netro
5	78	11.4	257	21	Arabidopsis thalia
6	78	11.4	258	21	Arabidopsis thalia
7	76.5	11.1	632	18	H. pylori ORF 02ge
8	76	11.1	736	22	Drosophila melanog
9	75.5	11.0	632	22	Helicobacter pylor
10	75	10.9	309	22	S. epidermidis ope
11	74.5	10.8	344	22	Putative P. abyssi

12	74	10.8	540	22	AAU56042	Propionibacterium
13	74	10.8	622	22	AB48832	Trametes hirsuta p
14	73.5	10.7	416	22	AAU30706	Novel human secret
15	72.5	10.6	478	20	AAW94764	Rat type 2 methion
16	72.5	10.6	790	22	AAU57583	Propionibacterium
17	72	10.5	327	20	AAW78247	Fragment of human
18	72	10.5	716	22	AAU12219	Human PRO4304 poly
19	72	10.5	773	21	AAAB15495	Human MEG-4 DNA.
20	71.5	10.4	1241	20	AA42167	Human nephrin prot
21	71.5	10.4	1241	22	AB47047	Human nephrin. Ho
22	71.5	10.4	1558	22	ABG28804	Novel human diago
23	70.5	10.3	478	20	AAW99599	Human methionine a
24	70.5	10.3	478	20	AAW93215	Human p67 homologu
25	70.5	10.3	478	20	AAW94763	Mouse type 2 methi
26	70.5	10.3	478	20	AAW94765	Human type 2 methi
27	70.5	10.3	478	22	ABB50275	eIF-2-associated p
28	70.5	10.3	478	22	AB28377	Human methionine a
29	70.5	10.3	500	22	AAW73991	Human colon cancer
30	69.5	10.1	552	18	AAW14287	Human deleted in p
31	69.5	10.1	552	21	AAW70071	Human tumour suppr
32	69.5	10.1	552	21	AAW69622	Human Smad4. Homo
33	69.5	10.1	798	19	AAW85025	Smad4-green floure
34	69.5	10.1	806	19	AAW85013	Smad4-green floure
35	69	10.0	1248	19	AAW69594	Filamentous haemag
36	69	10.0	1248	20	AAW43569	Amino acid sequenc
37	69	10.0	1248	20	AAW23960	Fragment 7 of the
38	69	10.0	1248	21	AAW67527	FHA fragment 7 pol
39	69	10.0	3596	21	AAW87407	Bordetella pertuss
40	69	10.0	3647	11	AAW05041	Filamentous haemag
41	68.5	10.0	231	21	AAW53334	Human colon cancer
42	68.5	10.0	1063	19	AAW59277	Rubella virus RA27
43	68.5	10.0	1134	21	AAW02005	Adenyl cyclase t
44	68	9.9	623	20	AAW94308	Coriolus versicolo
45	68	9.9	1252	21	AAW43197	Human ORF2961

#### ALIGNMENTS

RESULT 1

AA44650

ID AA44650 standard; Protein: 135 AA.

XX AC AA44650;

XX DT 18-APR-2000 (first entry)

XX DE Streptomyces albus G SsgA protein.

XX KW SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.

XX OS Streptomyces albus G.

XX PN WO200000613-A1.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99WO-NL00395.

XX PR 26-JUN-1998; 98EP-0202148.

XX PA (UYLE-) RIJKSUNIV LEIDEN.

XX PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.

XX PI Van Wezel GP, Kraal B, Luiten RGM;

XX DR WPI; 2000-147269/13.

XX DR N-PSDB; AA249729.

PT Reducing branching and enhancing fragmentation in filamentous  
PT microorganisms used to improve their liquid culturing properties -  
XX  
PS Disclosure; Fig 5; 60pp; English.  
XX  
CC The present sequence is S. albus G SsgA protein. SsgA reduces branching  
CC and fragment septation and enhances fragmentation of mycelium in liquid  
CC culture resulting in lower viscosity of culture broths. Filamentous  
CC bacteria can be transformed with ssgA gene-containing plasmid to enhance  
CC the production of secondary metabolites such as, antibiotics, antitumour  
CC agents, immunosuppressive agents, hypocholesterolaemic agents, enzyme  
CC inhibitors, antimigraine agents, herbicides, antiparasitic agents,  
CC ruminant growth promoters, bioinsecticides, receptor agonists and  
XX antagonists and biomass.  
XX  
SQ Sequence 135 AA;  
  
Query Match 100.0%; Score 687; DB 21; Length 135;  
Best Local Similarity 100.0%; Pred. No. 4.le-75;  
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 MSFLVSEELAFRIPVELRYETVDPYAVRLTFLHPLGDAPVTWVFGRELVEGVLDAGDGD 60  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 msflvseelafrripvelryetvdpvavrltflhplgdapvtwvfgrellivegldagdg 60  
  
Qy 61 VRVCPVGQTATREVHITLQVSEQALFRVGKAPLLAFLDRTDQGLSGSERAHADFDShL 120  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 61 vrvcpgqtatrevhltqvgseqalfrvgkapllafldrtdqglsgsgerahadfdshl 120  
  
Qy 121 DDALNRSIAEQSAG 135  
|||||:|||||  
Db 121 ddalnrsiaeqsag 135  
  
RESULT 2  
AAY44651  
ID AAY44651 standard; Protein; 135 AA.  
XX  
AC AAY44651;  
XX  
DT 18-APR-2000 (first entry)  
DE Streptomyces goldeniensis SsgA protein.  
KW SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.  
XX  
OS Streptomyces goldeniensis.  
XX  
PN WO200000613-A1.  
XX  
PD 06-JAN-2000.  
XX  
PF 25-JUN-1999; 99WO-NL00395.  
XX  
PR 26-JUN-1998; 98EP-0202148.  
XX  
PA (UYLE-) RIJKSUNIV LEIDEN.  
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.  
XX  
PI Van Wezel GP, Kraal B, Luiten RGM;  
XX  
DR WPI; 2000-147269/13.  
XX  
PS Disclosure; Fig 5; 60pp; English.

XX The present sequence is S. goldeniensis SsgA protein. SsgA reduces  
CC branching and fragment septation and enhances fragmentation of mycelium  
CC in liquid culture resulting in lower viscosity of culture broths.  
CC Filamentous bacteria can be transformed with ssgA gene-containing  
CC plasmid to enhance the production of secondary metabolites such as,  
CC antibiotics, antitumour agents, immunosuppressive agents,  
CC hypocholesterolaemic agents, enzyme inhibitors, antimigraine agents,  
CC herbicides, antiparasitic agents, ruminant growth promoters,  
CC bioinsecticides, receptor agonists and antagonists and biomass.  
XX  
SQ Sequence 135 AA;  
  
Query Match 70.0%; Score 481; DB 21; Length 135;  
Best Local Similarity 71.18; Pred. No. 3.6e-50;  
Matches 96; Conservative 14; Mismatches 25; Indels 0; Gaps 0;  
  
Qy 1 MSFLVSEELAFRIPVELRYETVDPYAVRLTFLHPLGDAPVTWVFGRELVEGVLDAGDGD 60  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 msflvseelafrripvelryetvdpvavrltflhplgdapvtwvfgrellidagprpcgdgd 60  
  
Qy 61 VRVCPVGQTATREVHITLQVSEQALFRVGKAPLLAFLDRTDQGLSGSERAHADFDShL 120  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 61 vhiapadpetfgevlrlqvgseqalfrvgkapllafldrtdkivplggersladfali 120  
  
Qy 121 DDALNRSIAEQSAG 135  
|||||:|||||:|||||  
Db 121 dealdnrslaeqsnag 135  
  
RESULT 3  
AAY44649  
ID AAY44649 standard; Protein; 135 AA.  
XX  
AC AAY44649;  
XX  
DT 18-APR-2000 (first entry)  
DE Streptomyces griseus SsgA protein.  
XX  
KW SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.  
XX  
OS Streptomyces griseus.  
XX  
PN WO200000613-A1.  
XX  
PD 06-JAN-2000.  
XX  
PF 25-JUN-1999; 99WO-NL00395.  
XX  
PR 26-JUN-1998; 98EP-0202148.  
XX  
PA (UYLE-) RIJKSUNIV LEIDEN.  
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.  
XX  
PI Van Wezel GP, Kraal B, Luiten RGM;  
XX  
DR WPI; 2000-147269/13.  
XX  
PS Disclosure; Fig 5; 60pp; English.  
XX  
PT Reducing branching and enhancing fragmentation in filamentous  
PT microorganisms used to improve their liquid culturing properties -  
XX  
PS Disclosure; Fig 5; 60pp; English.

CC bacteria can be transformed with ssrA gene-containing plasmid to enhance  
CC the production of secondary metabolites such as, antibiotics, antitumour  
CC agents, immunosuppressive agents, hypocholesterolaemic agents, enzyme  
CC inhibitors, antimigraine agents, herbicides, antiparasitic agents,  
CC ruminant growth promoters, bioinsecticides, receptor agonists and  
CC antagonists and biomass.  
XX  
SQ Sequence 135 AA;

Query Match 67.5%; Score 464; DB 21; Length 135;  
Best Local Similarity 65.9%; Pred. No. 4.1e-48;  
Matches 89; Conservative 16; Mismatches 28; Indels 0; Gaps 0;

Qy 1 MSFLVSEELAPRIPIVELRYETVDYAVRLTFHLPDGPVTVWFGRELLVEGVLDAGDGD 60  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 msflvseelsrlpvelryevgdpayalrmtfhlpgdapvtwafgrellldginspsgdgd 60  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|  
Qy 61 VRVCPVGQTATREYHITLVGSEQALFRVGKAPLLAFLDRTDQGLSGSERAHADFDShL 120  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|  
Db 61 vhiqtepegldvhiriqvadralfraqtaplvafldrtkivpgqehltdfgnll 120  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|  
Qy 121 DDALNRSIAEQSAG 135  
:|||||:|||||:|  
Db 121 edalgrillaeeqnag 135

RESULT 4  
AAI44652  
ID AAY44652 standard; Protein; 135 AA.

XX AC AAY44652;

XX DT 18-APR-2000 (first entry)

XX DE Streptomyces netropsis SsgA protein.

XX KW SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic; antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.  
XX

OS Streptomyces netropsis.

XX PN WO200000613-A1.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99WO-NL00395.

XX PR 26-JUN-1998; 98EP-0202148.

XX PA (UYLE-) RIJKSUNIV LEIDEN.

XX PA (NEW-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.

XX PI Van Wezel GP, Kraal B, Luiten RGM;

XX DR WPI; 2000-147269/13.

XX DR N-PSDB; AAZ49731.

PT Reducing branching and enhancing fragmentation in filamentous  
PT microorganisms used to improve their liquid culturing properties -

PS Disclosure; Fig 5; 60pp; English.

XX CC The present sequence is S. netropsis SsgA protein. SsgA reduces  
CC branching and fragment septation and enhances fragmentation of mycelium  
CC in liquid culture resulting in lower viscosity of culture broths.  
CC Filamentous bacteria can be transformed with ssrA gene-containing  
CC plasmid to enhance the production of secondary metabolites such as,  
CC antibiotics, antitumour agents, immunosuppressive agents,  
CC hypocholesterolaemic agents, enzyme inhibitors, antimigraine agents,

CC herbicides, antiparasitic agents, ruminant growth promoters,  
CC bioinsecticides, receptor agonists and antagonists and biomass.  
XX  
SQ Sequence 135 AA;

Query Match 64.9%; Score 446; DB 21; Length 135;  
Best Local Similarity 63.0%; Pred. No. 6.2e-46;  
Matches 85; Conservative 20; Mismatches 30; Indels 0; Gaps 0;

Qy 1 MSFLVSEELAPRIPIVELRYETVDYAVRLTFHLPDGPVTVWFGRELLVEGVLDAGDGD 60  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 msflvseelskplvelryetrdpyavmtfhlpgdapvtwafgrellldginspsgdgd 60  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

Qy 61 VRVCPVGQTATREYHITLVGSEQALFRVGKAPLLAFLDRTDQGLSGSERAHADFDShL 120  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|  
Db 61 vhiaptdeglsdsvirlqvadralfraqtaplvafldrtkdvspigqetlgdfedsl 120  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|  
Qy 121 DDALNRSIAEQSAG 135  
:|||||:|||||:|  
Db 121 eaalgrillaeeqnag 135

RESULT 5  
AAG35780  
ID AAG35780 standard; Protein; 257 AA.

XX AC AAG35780;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 43758.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 08-APR-1999; 99US-0128234.

XX PR 16-APR-1999; 99US-0128714.

XX PR 19-APR-1999; 99US-0129845.

XX PR 21-APR-1999; 99US-0130077.

XX PR 23-APR-1999; 99US-0130449.

XX PR 28-APR-1999; 99US-0130510.

XX PR 30-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 04-MAY-1999; 99US-0132407.

XX PR 05-MAY-1999; 99US-0132484.

XX PR 06-MAY-1999; 99US-0132485.

XX PR 07-MAY-1999; 99US-0132487.

XX PR 11-MAY-1999; 99US-0132863.

XX PR 14-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 18-MAY-1999; 99US-0134370.

XX PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0156559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.

```
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      11.4%; Score 78; DB 21; Length 257;
Best Local Similarity 29.7%; Pred. No. 0.53;
Matches 30; Conservative 14; Mismatches 35; Indels 22; Gaps 4;

Qy 50 EGVLDAAAGDGVRCVPGQTATREVIHTLVGSEQALFRVGKAPLLAFDRDQGLSLGS 109
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39 dgtsdsdpppk-pegdtrrqellari-----amigtkskvrldfiderseyltkfa 91

Qy 110 ERAHADFD-----SHLDDALNRSI-----AEEQSAG 135
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 92 eanaefdkvgedamkdideastrilenieskmqafeesag 132

RESULT 6
AAG35779
ID AAG35779 standard; Protein; 258 AA.
XX
AC AAG35779;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 43757.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134769.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
```



CC were analysed for significant homology to other known or exported  
CC membrane proteins. Having identified and determined the sequences of  
CC interest, particular regions can be isolated from H. pylori by PCR  
CC amplification for recombinant polypeptide production, e.g. in E. coli  
CC hosts.

XX SQ Sequence 632 AA;

Query Match 11.1%; Score 76.5; DB 18; Length 632;

Best Local Similarity 29.9%; Pred. No. 2.8;

Matches 29; Conservative 15; Mismatches 42; Indels 11; Gaps 4;

QY 44 GRELLVEGVLDAGDGVPCVGTATREVTHTLVQVSEQALFRVKG--APLLAFLDRT 101

Db 215 gkllakav---ageahvffsmggssfiemfvglqasrvrdlftakqapsiifidei 271

QY 102 DQGLSLGSRERAHADFSLDD---ALNRSLEAEQSAG 135

Db 272 d---aigkaraaggmisdndereqtlqlaemdggf 305

RESULT 8

ABB59950

ID ABB59950 standard; Protein; 736 AA.

XX AC ABB59950;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 6642.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL04053.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -

XX PS Disclosure; SEQ ID NO 6642; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 736 AA;

Query Match 11.1%; Score 76; DB 22; Length 736;

Best Local Similarity 23.3%; Pred. No. 3.9;

Matches 40; Conservative 24; Mismatches 62; Indels 46; Gaps 7;

QY 3 FLVSEELAFRIPVELRYETVDPYAVRLTF-----HLPG 35

Db 272 fttsgsvfrliqlgnqve-vdpeeinvtfedvkgcdeakqlkevveflkspekfsnlg 330

QY 36 DAPVTWVF-----GRELLVEGVLDAGDGVPCVGTATREVTHTLVQVSEQALFRV 89

Db 331 klpkgvllvpppgtgktilarav---ageakvpffhaagpefdevivggarrvrdlifa 387

QY 90 GK--APLLAFLDRTDQGLSLGSRERAHADFSLDDALNRSLEAE---EQSAG 135

Db 388 akarapcvifideid---svgaktntsvlhpyanqinqlisemdghfnag 436

RESULT 9

AAU35800

ID AAU35800 standard; Protein; 632 AA.

XX AC AAU35800;

XX DT 14-FEB-2002 (first entry)

XX DE Helicobacter pylori cellular proliferation protein #113.

XX KW Antisense; prokaryotic cellular proliferation protein;

XX KW antibiotic; antibacterial; drug design.

XX OS Helicobacter pylori.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207727P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlisen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX DR N-PSDB; AAS53659.

XX PT New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -

XX PS Example 3; Seq ID No 11393; 511pp; English.

XX CC The invention relates to antisense inhibitors of genes essential to  
XX prokaryotic cellular proliferation, their use in identifying the  
XX genes, their use in the discovery of novel antibiotics, the essential  
XX genes themselves and the encoded proteins. The prokaryotes used are  
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
XX invention is also useful for the identification of potential new targets  
XX for antibiotic development. The antisense nucleic acids can also be used  
XX to identify proteins used in proliferation, to express these proteins,  
XX and to obtain antibodies capable of binding to the expressed proteins.

XX CC The proteins can be used to screen compounds in rational drug discovery  
XX programmes. The antisense nucleic acid sequence is also useful to screen  
XX for homologous nucleic acids which are required for cell proliferation in





Matches 39; Conservative 18; Mismatches 51; Indels 35; Gaps 6;

QY 2 SFVLSEELAFRIPVELRYETVDPYAVRLTFHLPGDAPVTWFGRELLVEGV-----LDAA 56  
||| : ||| | : | | | | | : ||| | :  
Db 190 sfeeaislgrvgvmlrglrivqgvdkvefnsdpdegvarflgfenilegvakgnlleean 249

QY 57 G-----DGDVRVC-----PVGQTATREVHITLQVSEQ-----ALFRVG 90  
| : ||| | : | | | : | : | :  
Db 250 gvkitpisvegkrigvrpedilistepvktsarnefraev-igieelglprlvnllkg 308

QY 91 KAPLLAFLDRT----DOGLSLGSE 110  
| ||| : | : | | | |  
Db 309 gitlkafirsslieligsegre 331

RESULT 12  
AAU56042  
ID AAU56042 standard; Protein; 540 AA.  
XX AC AAU56042;  
XX DT 27-FEB-2002 (first entry)  
XX DE Propionibacterium acnes immunogenic protein #16938.  
XX OS SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX XX  
OS Propionibacterium acnes.  
XX WO200181581-A2.  
XX PN  
XX PD 01-NOV-2001.  
XX PF 20-APR-2001; 2001WO-US12865.  
XX PR 21-APR-2000; 2000US-199047P.  
XX PR 02-JUN-2000; 2000US-208841P.  
XX PR 07-JUL-2000; 2000US-216747P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
XX DR N-PSDB; AAS59573.  
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX XX  
Claim 3; SEQ ID No 17237; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 540 AA;

Query Match 10.8%; Score 74; DB 22; Length 540;  
Best Local Similarity 25.8%; Pred. No. 4.5;  
Matches 32; Conservative 17; Mismatches 51; Indels 24; Gaps

QY 16 ELRYETVDPEYAVRLTFHLPGDA--PVTWFGRELLVEGVLDAGDGYRVCPVGQTATRE 73  
::| | | | | : | | | : | | | : | | | :  
Db 211 qiyystadgvderdvelptdahtmswafskpgykvafaa-----tlstpqngasfga 265

QY 74 VHTLVQVSEQALFRVKAPLLAFLDRT--DOGLSGSERAHADFDSDALNLSLAEE 131  
:| : ||| : | | | | | | | | | | | : | :  
Db 266 qltiavgadprtii-----pela--drvtvdgg-----hadlsadidegtmtildsp 310

QY 132 QSAG 135  
|  
Db 311 tggg 314

RESULT 13  
AAB48832  
ID AAB48832 standard; Protein; 622 AA.  
XX AC AAB48832;  
XX DT 13-MAR-2001 (first entry)  
XX DE Trametes hirsuta pyranose oxidase.  
XX KW Pyranose oxidase; expression construct; recombinant production;  
KW monosaccharide oxidation; 2-keto derivative;  
KW hydrogen peroxide production.  
XX OS Trametes hirsuta.  
XX PN US6146865-A.  
XX PD 14-NOV-2000.  
XX PF 05-MAY-1999; 99US-0305381.  
XX PR 08-JUN-1998; 98DK-0000774.  
XX PR 10-JUN-1998; 98US-0088724.  
XX PA (NOVO ) NOVO NORDISK AS.  
XX PI Schneider P, Christensen S, Lassen SF;  
XX WPI; 2001-049055/06.  
XX DR N-PSDB; AAC87518, AAC87519.  
XX PT Novel nucleic acid molecule encoding polypeptide having pyranose  
PT oxidase activity used to design oligonucleotide probes to identify and  
PT clone DNA encoding the polypeptide from different genera or species -  
XX  
Claim 5; Fig 1; 20pp; English.

The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also relates to expression constructs, expression vectors and recombinant cells comprising pyranose oxidase nucleic acid sequences, and the recombinant production of Trametes hirsuta pyranose oxidase. Pyranose oxidase catalyses the oxidation of several monosaccharides in the pyranose form at position C2 to produce 2-keto derivatives with the release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta pyranose oxidase may be used to produce the enzyme and to design oligonucleotide probes to identify and clone genomic pyranose oxidase cDNA or genomic DNA from different genera or species of microorganisms

CC (fungi or bacteria). The present sequence represents pyranose oxidase  
CC from the fungus Trametes hirsuta.

SQ Sequence 622 AA;

Query Match 10.8%; Score 74; DB 22; Length 622;

Best Local Similarity 26.2%; Pred. No. 5.4;

Matches 28; Conservative 17; Mismatches 32; Indels 30; Gaps 6;

QY 26 AVRLTHL-----PGDAPVTWVFGRELLVEGLDAGGDVRCVPGQTATREVHITLQ 79

Db 22 akatahsipplpgpdlp-----pgmnevdyalvgsg-----pigctyarel---ve 67

QY 80 VGSEQALFRVGKAPLLAFLDRTDQGLSLGSERAH-ADFSDHLDDALN 125

Db 68 agfnvamfeige-----idsgkigshkntvayqknidkfvn 105

RESULT 14

AAU30706

ID AAU30706 standard; Protein; 416 AA.

XX

AC AAU30706;

XX

DT 18-DEC-2001 (first entry)

XX

DE Novel human secreted protein #1197.

XX

KW Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200179449-A2.

XX

PD 25-OCT-2001.

XX

PF 16-APR-2001; 2001WO-US08656.

XX

PR 18-APR-2000; 2000US-0552929.

XX

PR 26-JAN-2001; 2001US-0770160.

XX

PA (HYSE-) HYSEQ-INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-611725/70.

XX

PT Nucleic acids encoding a range of human polypeptides, useful in genetic

PT vaccination, testing and therapy -

XX

PS Claim 20; Page 333; 765pp; English.

XX

CC The invention relates to novel human secreted polypeptides. The  
CC polypeptides and antibodies to the polypeptides are useful for  
CC determining the presence of or predisposition to a disease associated  
CC with altered levels of polypeptide. The polypeptides are also useful for  
CC identifying agents (agonists and antagonists) that bind to them. Cells  
CC expressing the proteins are useful for identifying a therapeutic agent  
CC for use in treatment of a pathology related to aberrant expression or  
CC physiological interactions of the polypeptide. Vectors comprising  
CC the nucleic acids encoding the polypeptides and cells genetically  
CC engineered to express them are also useful for producing the proteins.  
CC The proteins are useful in genetic vaccination, testing and  
CC therapy, and can be used as nutritional supplements. They may be used to  
CC increase stem cell proliferation; to regulate haematopoiesis; and in  
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid  
CC sequences of novel human secreted proteins of the invention.

XX

SQ Sequence 416 AA;

Query Match 10.7%; Score 73.5; DB 22; Length 416;

Best Local Similarity 27.0%; Pred. No. 3.6;

Matches 33; Conservative 18; Mismatches 40; Indels 31; Gaps 6;

QY 14 PYELRYETVDPYAVR-----LTFHLPGDAPVTWVFGRELLVE-----GVLDAAQ 57

Db 133. pvliteaplnpkankrekmtqllcfetfntpghvp--wpiqavlsxlswaqpigivmdsg 190

QY 58 DGDVRCVPVGGQTATREVHITLQVGSQALFRVGKAPLLAFLDRTDQGLSLGSERAHADF 117

Db 191. dgvtvcvlpilrgattllhailrlg-----pgla-rdltdylmkilttergys-ft 237

QY 118 SH 119

Db 238 th 239

RESULT 15

AAW94764

ID AAW94764 standard; protein; 478 AA.

XX

AC AAW94764;

XX

DT 28-APR-1999 (first entry)

XX

DE Rat type 2 methionine aminopeptidase (MetAP2).

XX

KW Ovalicin; fumagillin; type 2 methionine aminopeptidase; inhibitor;

KW MetAP2; angiogenesis; tumour; diabetic retinopathy; arteriosclerosis;

KW inflammatory disease; immune reaction; autoimmune disease; allergy;

KW tissue graft rejection; rat.

XX

OS Rattus sp.

XX

PN WO9856372-A1.

XX

PD 17-DEC-1998.

XX

PF 08-JUN-1998; 98WO-US11775.

XX

PR 09-JUN-1997; 97US-0049159.

XX

PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.

XX

PI Griffith EC, Liu JO, Su Z;

XX

DR WPI; 1999-080848/07.

XX

PT New ovalicin and fumagillin derivatives - are inhibitors of type 2

PT methionine amino-peptidase, useful for treating or diagnosing

PT diseases involving abnormal angiogenesis or immune reactions

XX

PS Disclosure; Fig 2; 99pp; English.

XX

CC The invention relates to ovalicin and fumagillin derivatives that can  
CC inhibit type 2 methionine aminopeptidase (MetAP2). These inhibitors are  
CC useful for treating and/or diagnosing diseases involving abnormal  
CC angiogenesis (e.g. tumours, diabetic retinopathy, inflammatory diseases,  
CC arteriosclerosis) or immune reactions which result in pathology (e.g.  
CC autoimmune disease, allergy and tissue graft rejection). The present  
CC sequence represents the amino acid sequence of rat MetAP2.

SQ Sequence 478 AA;

Query Match 10.6%; Score 72.5; DB 20; Length 478;

Best Local Similarity 28.8%; Pred. No. 5.7;

Matches 32; Conservative 20; Mismatches 34; Indels 25; Gaps 8;

QY 17 LRYETVDPYAVRLTF--HLPG---DAPVTWVFG--RELLVEGLDAGDG-----DVRV 63

Db 242 lqyddi-----ckidfgthisgriidcaftvtfnpkydillkavkdatntgikcagidvrl 297

Qy 64 CPVGTATREVHITLQVSGEQALFRVGKAPLLAFLDRTDQGLSLGSERAHA 114

Db 298 cdvge-aiqevmesyeveidgktyqv--kpi-----rnlnghsigpyriha 340

Search completed: July 18, 2002, 14:15:02

Job time: 10332 sec

**THIS PAGE BLANK (USPTO)**



RESULT 4  
US-09-040-

, PATENT NO.: 5083620  
 , GENERAL INFORMATION:  
 , APPLICANT: CHANG, YIE-HWA  
 , TITLE OF INVENTION: CLONE OF A NUCLEOTIDE SEQUENCE ENCODING  
 , TITLE OF INVENTION: A PROTEIN HAVING TWO FUNCTIONS  
 , NUMBER OF SEQUENCES: 5  
 , CORRESPONDENCE ADDRESS:  
 , ADDRESSEE: HOWELL & HAFFERKAMP, L.C.  
 , STREET: 7733 FORSYTH BLVD., SUITE 1400  
 , CITY: ST. LOUIS  
 , STATE: MO  
 , COUNTRY: USA  
 , ZIP: 63105

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/040,799  
 FILING DATE: 18-MAR-1998  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: HOLLAND, DONALD R  
 REGISTRATION NUMBER: 35,197  
 REFERENCE/DOCKET NUMBER: 16153-4639  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 314-727-5188  
 TELEFAX: 314-727-6092  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:

```

; 10 amino acids
; 20 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-040-799-3

Query Match          10.3%   Score 70.5;   DB 2;   Length 478;
Best Local Similarity 28.8%;   Pred. No. 2,3;
Matches 32;   Conservative 19;   Mismatches 35;   Indels 25;   Gaps 8;

QY  17  LRYETVDPYAVRLTF--HLPG---DAPYTWFG--RELLVEGYLDAAGD-----DYRV 63
      |||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db   242  LQYDDI----CKIDFGHISGRIDCAFTFNPKYDFLLKAVKDADTWTGKIVACIDVRL 297

QY  64  CPVGQTATREVHITLQVGEQALFRVCGKAPLLAFLDRTDQGLSLGSERAHA 114
      |||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db   298  CDVGE-AIQEVMSVEVEIDGKTKYQV--KPI-----RNLNGHSIGQYRIHA 340

```

```

; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 0492611-0346

```

; CURRENT APPLICATION NUMBER: US/09/093.448A  
; CURRENT FILING DATE: 1998-06-08  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 478  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-093-448-1

Query Match 10.3%; Score 70.5; DB 4; Length 478;  
Best Local Similarity 28.8%; Pred. No. 2.3;  
Matches 32; Conservative 19; Mismatches 35; Indels 25; Gaps 8;

Qy 17 LRYETVDPYAVRLTF--HLPG---DAPVTWVFG--RELLVGVLDAAAGDG-----DVRV 63  
Db 242 LOYDDI-----CKIDFGTHISGRIDCAFTVTNPKYDILLKAVKDATWTGKACIDVRL 297  
Qy 64 CPVGGTATREVHITLQVGEALFRVGRKAPLLAFLDRTDQGLSGSRAHA 114  
Db 298 CDVGE-ALQEVWESVEIDKTKYQV--KPI-----RNLNGHSIGPYRIHA 340

RESULT 6  
US-09-093-448-3  
; Sequence 3, Application US/09093448A  
; Patent No. 6207704  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jun O.  
; APPLICANT: Griffith, Eric C.  
; APPLICANT: Su, Zhuang  
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors  
; FILE REFERENCE: 0492611-0346  
; CURRENT APPLICATION NUMBER: US/09/093.448A  
; CURRENT FILING DATE: 1998-06-08  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 478  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-093-448-3

Query Match 10.3%; Score 70.5; DB 4; Length 478;  
Best Local Similarity 28.8%; Pred. No. 2.3;  
Matches 32; Conservative 19; Mismatches 35; Indels 25; Gaps 8;

Qy 17 LRYETVDPYAVRLTF--HLPG---DAPVTWVFG--RELLVGVLDAAAGDG-----DVRV 63  
Db 242 LOYDDI-----CKIDFGTHISGRIDCAFTVTNPKYDILLKAVKDATWTGKACIDVRL 297  
Qy 64 CPVGGTATREVHITLQVGEALFRVGRKAPLLAFLDRTDQGLSGSRAHA 114  
Db 298 CDVGE-ALQEVWESVEIDKTKYQV--KPI-----RNLNGHSIGPYRIHA 340

RESULT 7  
US-08-588-821-2  
; Sequence 2, Application US/08588821  
; Patent No. 5712097  
; GENERAL INFORMATION:  
; APPLICANT: Kern, Scott E.  
; APPLICANT: Hahn, Stephan A.  
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, DPC4  
; NUMBER OF SEQUENCES: 91  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA

; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/588.821  
; FILING DATE: 19-JAN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07265/079001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 552 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-588-821-2

Query Match 10.1%; Score 69.5; DB 1; Length 552;  
Best Local Similarity 30.9%; Pred. No. 3.7;  
Matches 30; Conservative 15; Mismatches 27; Indels 25; Gaps 8;

Qy 5 VSEELAFRIPV-----ELRYETVDPYAVRLTFHLPGDAPVTWVFGRELLVGVLD 54  
Db 304 VHNELAFQPPISNHPAPEYWCIAFYEMD--VOVGETFRVPSCPPI-----VTVDGYVD 355  
Qy 55 AAGDGDVRVCPVGGTATREVHITLQVGEALFRVGRK 91  
Db 356 PSG-GD-RFC-LGQLS--NVHRTETAI--ERARLHICK 385

RESULT 8  
US-08-915-214-2  
; Sequence 2, Application US/08915214  
; Patent No. 5814457  
; GENERAL INFORMATION:  
; APPLICANT: Kern, Scott E.  
; APPLICANT: Hahn, Stephan A.  
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, DPC4  
; NUMBER OF SEQUENCES: 91  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/915.214  
; FILING DATE: 20-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/588.821  
; FILING DATE: 19-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07265/079001  
; TELECOMMUNICATION INFORMATION:

```

Db      304 VHNELAFQPPISNHPAPEYWCSTAYEMD-VQGETFKVPSSCPI-----VTVDGYVD 355
Qy      55 AAGGDGVRCVPGQTATREVHITLQVGSEALFRVK 91
Db      356 PSG-GD-RFC-LQLS--NVHRTEAI--ERARLHIK 385

RESULT 10
US-08-701-582D-14
; Sequence 14, Application US/08701582D
; Patent No. 6017755
; GENERAL INFORMATION:
; APPLICANT: WRANA, Jeffrey
; APPLICANT: ATTISANO, Liliana
; APPLICANT: SCHERER, Stephen W.
; TITLE OF INVENTION: MADR2 TUMOUR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/701,582D
; FILING DATE: 22-Aug-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REA, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 024916-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-701-582D-14

Query Match          10.1%; Score 69.5; DB 3; Length 552;
Best Local Similarity 30.9%; Pred.No. 3.7;
Matches 30; Conservative 15; Mismatches 27; Indels 25; Gaps

Qy      5 VSELEAPRIPV-----ELRYETVDPYAVRLTFHLPGDAPVTVWFGRELLVEGLVD 54
Db      304 VHNELAFQPPISNHPAPEYWCSTAYEMD-VQGETFKVPSSCPI-----VTVDGYVD 355
Qy      55 AAGGDGVRCVPGQTATREVHITLQVGSEALFRVK 91
Db      356 PSG-GD-RFC-LQLS--NVHRTEAI--ERARLHIK 385

RESULT 11
US-09-096-776B-9
; Sequence 9, Application US/09096776B
; Patent No. 627094
; GENERAL INFORMATION:
; APPLICANT: Miyazono, Kohei
; APPLICANT: Kawabata, Masahiro
; TITLE OF INVENTION: SNAD6 AND USES THEREOF
; FILE REFERENCE: L0461/7038

```



```

; CURRENT APPLICATION NUMBER: US/09/096,776B
; CURRENT FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: US 60/049,990
; PRIOR FILING DATE: 1997-06-13
; PRIOR APPLICATION NUMBER: US 60/053,040
; PRIOR FILING DATE: 1997-07-18
; PRIOR APPLICATION NUMBER: US 60/066,173
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-096-776B-9

```

Query Match 10.1%; Score 69.5; DB 4; Length 552;  
Best Local Similarity 30.9%; Pred. No. 3.7;

Qy 5 VSEELAFRIPV-----ELRYETVDPYAVRLTETHLPGDAPVTWVFGRELLVEGVLD 54  
+ + + + + : : : : + + + + : : : :  
Db 304 VNELAFQDPISNHPAPEYWCISIAFYEMD-VQVGETEKVPSSCP I-----VTVDGYVD 355

QY 55 AAGDGDVRVCPVGTATREHVHITLQVGSEQALFRVGK 91  
:| | | | | : | | : | | :| |  
Db 356 PSG-GD-RFC-LGQLS--NVHRTEAI--ERARLHIGK 388

```

RESULT 12
US-08-348-353-17
: Sequence 17, Application US/08348353
: Patent No. 5932217
: GENERAL INFORMATION:
: APPLICANT: Tuomanen, Elaine
: APPLICANT: Masure, Robert
: TITLE OF INVENTION: Antibody Recognizing Endothelial Cell
: TITLE OF INVENTION: Ligand for Leukocyte CR3
: NUMBER OF SEQUENCES: 38
: CORRESPONDENCE ADDRESS:
:

```

/ CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Klauber & Jackson  
 / STREET: 411 Hackensack Avenue  
 / CITY: Hackensack  
 / STATE: New Jersey  
 / COUNTRY: U.S.A.  
 / ZIP: 07601  
 /

```

:
:
: COMPUTER READABLE FORM:
:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.25
:
:   CURRENT APPLICATION DATA:
:
:   APPLICATION NUMBER: US/08/348,353
:   FILING DATE: 30-NOV-1994
:   CLASSIFICATION: 514
:   ATTORNEY/AGENT INFORMATION:
:
:
:

```

```

, NAME: Jackson, David
, REGISTRATION NUMBER: 26,742
, REFERENCE/DOCKET NUMBER: 600-1-097C1P1
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 201-487-5800
, TELEFAX: 201-343-1684
, INFORMATION FOR SEQ ID NO: 17:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1248 amino acids
, TYPE: amino acid
, TOPOLOGY: linear
, MOLECULE TYPE: protein
,

```

Query Match 10.0%; Score 69; DB 2; Length 1248;

	Best Local Similarity	28.18; Pred. No. 14;							
	Matches	Conservative	10; Mismatches	37; Indels	22; Gaps				
Qy	47	LLVEGVLDAAGDGVDRVCPVGOTATREVHITLQVSGEQALFRVGRKAPLLAFLD-----	99						
Db	85	LAANGVADVNTGQDVRVAKL-----	VS	DAGADLQAGRSMTLGIVDTGDLQA	131				
Qy	100	RTDQGLSLGSRRAHADFDLSHLDLALNRSLAEQSG	135						
Db	132	RAQQLKLELGSVKSGDGLQAAAAGGAL--SLAAAEVAG	165						

RESULT 13  
US-08-465-965-17  
: Sequence 17, Application US/08465965  
: Patent No. 5968512  
: GENERAL INFORMATION:  
: APPLICANT: Tuomanen, Elaine  
: APPLICANT: Masure, Robert  
: TITLE OF INVENTION: Antibody Recognizing Endothelial Cell  
: TITLE OF INVENTION: Ligand for Leukocyte CR3  
: NUMBER OF SEQUENCES: 38  
: CORRESPONDENCE ADDRESS:

CONNECTIONS ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07601

```
,  
,  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
,
```

; APPLICATION NUMBER: US/08/465,965  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/348,353  
 ; FILING DATE: 30-NOV-1994

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/247,572

```

; FILING DATE: 23-MAY-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DCE/HUS92/03725

```

```

;
; APPLICATION NUMBER: PCT/US92/0
; FILING DATE: 04-MAY-1992
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

```

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/695,613  
 ; FILING DATE: 03-MAY-1991  
 ; CLASSIFICATION: 424

; CESSATION INFORMATION: 424  
 ;  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jackson, David  
 ; REGISTRATION NUMBER: 26,742  
 ;

REFERENCE/DOCKET NUMBER: 600-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

US-08-465-965-17

Query Match 10.0%; Score 69; DB 2; Length 1248;  
Best Local Similarity 28.1%; Pred. No. 14;  
Matches. 27; Conservative 10; Mismatches 37; Indels





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:17:33 ; Search time 73.98 Seconds  
(without alignments)  
175.345 Million cell updates/sec

Title: US-09-749-185-5

Perfect score: 687

Sequence: 1 MSFLVSEELAFRIPVELRYE.....FDSLDDALNRLAEQSQS 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	498	72.5	136	2 T37179	probable regulator
2	203	29.5	142	2 T36147	probable regulator
3	155	22.6	138	2 T35247	probable regulator
4	117	17.0	142	2 T35319	probable regulator
5	90.5	13.2	484	2 S68694	hexokinase (EC 2.7
6	82	11.9	741	2 A83271	hypothetical prote
7	79	11.5	862	2 T36380	probable large ATP
8	78.5	11.4	332	2 AE2438	thiamin monophosph
9	78.5	11.4	638	2 T47267	cell cycle protein
10	78	11.4	346	1 H69789	probable alcohol d
11	77	11.2	693	2 F96037	alpha-galactoside
12	76.5	11.1	291	2 H70678	hypothetical prote
13	76.5	11.1	632	2 D71941	ATP-dependent zinc
14	75.5	11.0	248	2 F97414	hypothetical prote
15	75.5	11.0	248	2 AB2632	hypothetical prote
16	75.5	11.0	632	2 E64653	cell division prot
17	74.5	10.8	344	2 G75203	abc transporter, A
18	74.5	10.8	536	2 A99708	2,3-dihydroxybenzo
19	74.5	10.8	536	2 E85558	membrane bound zin
20	74.5	10.8	645	2 G81315	precorrin-3B C17-m
21	74	10.8	253	2 AG97694	precorrin-3B C17-m
22	74	10.8	253	2 H29920	probable sugar upt
23	73.5	10.7	526	2 D96014	hypothetical prote
24	73	10.6	179	2 T36503	probable transcript
25	73	10.6	517	2 G83311	cell division prot
26	73	10.6	702	2 E64250	cell division prot
27	73	10.6	709	2 S73497	methionyl aminopep
28	72.5	10.6	480	2 A46702	conserved hypothet
29	72.5	10.6	509	2 H87685	

30 72.5 10.6 921 2 AE0332 conserved hypothet  
31 72 10.5 797 2 A96232 succinoglycan bios  
32 72 10.5 797 2 AD3054 hypothetical prote  
33 71.5 10.4 655 2 E64752 probable dihydroxy  
34 71.5 10.4 1241 2 T37190 nephrin - human  
35 71 10.3 322 2 G83766 quinone oxidoreduc  
36 71 10.3 406 2 AI0767 probable glycosylt  
37 71 10.3 514 2 B83360 hypothetical prote  
38 70.5 10.3 283 2 C84321 hypothetical prote  
39 70.5 10.3 303 2 AE2776 methyltransferase  
40 70.5 10.3 321 2 C97556 probable methyltra  
41 70.5 10.3 478 1 DPHUM2 methionyl aminopep  
42 70.5 10.3 536 1 SYECEB 2,3-dihydroxybenzo  
43 70.5 10.3 606 2 AD1219 diol dehydratase-r  
44 70.5 10.3 640 2 G96034 conserved hypothet  
45 70 10.2 596 2 A96710 hypothetical prote

#### ALIGNMENTS

RESULT 1

T37179

probable regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T37179

R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: Z21598

A:Accession: T37179

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-136 <SEE>

A:Cross-references: EMBL:AL096823; PIDN:CAB46964.1; GSPDB:GN00070; SCOEDB:SCQ11.09

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: ssgA; SCOEDB:SCQ11.09

Query Match 72.5%; Score 498; DB 2; Length 136;  
Best Local Similarity 75.6%; Pred. No. 1.le-43;  
Matches 102; Conservative 9; Mismatches 24; Indels 0; Gaps 0;

QY 1 MSFLVSEELAFRIPVELRYETVPYAVRLTFLPDCDAPVTWVFGRELLVEGVLDAGDGD 60

Db 2 MSFLVSEELAFRIPVELRYETVPYAVRLTFLPDCDAPVTWVFGRELLVEGVLDAGDGD 61

QY 61 VRVCPVGQTATREVIITLQVSEQUALFRVKGKAPLLAFDLRTDQGLSLGSERAHADFDShL 120

Db 62 VRTAPVEPEPLAEVLRLQVSDQALFRSSAAPLVAFDLRTDKLVPLGQEGALADFDShL 121

QY 121 DDALNRLAEQSQSAG 135

Db 122 DEALDRILAEQSQSAG 136

RESULT 2

T36147

probable regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000

C:Accession: T36147

R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: Z21598

A:Accession: T36147

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-142 <SEE>

A:Cross-references: EMBL:AL096852; PIDN:CAB51005.1; GSPDB:GN00070; SCOEDB:SCE19A.24

A:Experimental source: strain A3(2)

C:Genetics:



R; Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
: Nature 406, 959-964, 2000  
A; Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A; Reference number: AB2950; MUID:20437337  
A; Accession: A83271  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-741 <STO>  
A; Cross-references: GB:AE004724; GB:AE004091; NID:g9949083; PIDN:AA06372.1; GSPDB:GN001010  
A; Experimental source: strain PA01  
C; Genetics:  
A; Gene: PA2984

```

Query Match      11.9%; Score 82; DB 2; Length 741;
Best Local Similarity 32.7%; Pred. No. 2.8;
Matches 33; Conservative 8; Mismatches 32; Indels 28; Gaps 5;

Qy    26  AVRITFHLPGDAPVTWVFGRELLVEGVLDAAAGDDVRVCPVGQ-----TATREVHIIQLQV 80
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    477 ALLPALLPPSPPEW-----GRAEVRVDVGQGLAVLVRTREHVLYDS 521

Qy    81  GSTQALFRVCK---APLLAFEL-RTDQGLSLGSERAHADEF 117
       | : | | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    522 GARQGAFDMGERVVVPVLRLDLKRLDLGLLT-----SHADND 558

```

RESULT 7  
T36380  
probable large ATP-binding protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:accession: T36380  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, April 1999  
A:Reference number: Z21573  
A:Accession: T36380  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-862 <OLI>  
A:Cross-references: EMBL:AL049628; PTDN:CAB40869.1; GSPDB:GN00070; SCOEDB:SCE94.20  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCE94.20

Query Match	11.5%;	Score 79;	DB 2;	Length 862;
Best Local Similarity	31.7%;	Pred. No. 6.7;		
Matches	39;	Conservative	6;	Mismatches 40;
				Indels 38;
				Gaps 6;

RESULT 8  
AE2438  
thiamin monophosphate kinase [imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AE2438  
R:kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriouchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE2438  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-332 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA0766760.1; PID:g17134199; GSPDB:GN00179  
C:Genetics:  
A:Experimental source: strain PCC 7120  
C:Gene: alr5061  
C:Superfamily: conserved hypothetical protein MJ0640

	Query Match	11.4%	Score 78.5;	DB 2;	Length 332;
	Best Local Similarity	25.4%;	Pred. No. 2.5;		
	Matches 32; Conservative	14;	Mismatches 41;	Indels 39;	Gaps 5;
QY	24 PYAVRLTFHLPGDAPVTWVFGRELLVGEVLDAAG-----DGDVRVCPV---GQTATRE 73	: :               : :               :			
Dd	82' PLGTIALALPGDLDSVSWV---ERLYOGITCLOKQYHTPIVGSDVRSPITLSITAFGQ 138	: :               : :               :			
QY	74 VH-----ITLVQGSQAALFRVCVKAPLLAFLDRDTDOGLSGSERAHADFSHLDDALNR 126	: :               : :               :			
Dd	139 VHPNRIIRSTAQVGDAIATGVHCA-----SHAGLELLDPKIGK 179	: :               : :               :			
QY	127' SLAEQQ 132	: :               : :               :			
Dd	180' DLTPEE 185	: :               : :               :			

```

RESULT      9
T47267
cell cycle protein [imported] - Helicobacter felis
C:Species: Helicobacter felis
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C:Accession: T47267
R:Bayle, D.; Wangler, S.; Weitzenegger, T.; Steinhilber, W.; Volz, J.; Przybylski, M.
J. Bacteriol. 180, 317-329, 1998
A:Title: Properties of the P-type ATPases encoded by the copAP operons of Helicobacter
A:Reference number: Z24437; MUID: 98101471
A:Accession: T47267
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-638 <BAY>
A:Cross-references: EMBL: A001932; NID: g2660538; PIDN: CAA05102.1; PID: g2660540
A:Experimental source: strain ATCC 49179
C:Genetics:
A:Gene: ftsh
C:Superfamily: cell division protein ftsh; FtsH/SEC18/CDG48-typc ATP-binding domain

```

```

Query Match      11.4% ; Score 78.5 ; DB 2 ; Length 638 ;
Best Local Similarity 29.9% ; pred. No. 5.3 ;
Matches 29 ; Conservative 16 ; Mismatches 41 ; Indels 11 ; Gaps 4 ;

Qy 44 GRELLVEGVLDAAAGDGVRCVPVGTATREVIHTLVQVSEQALFRVGK--APLAFALDRT 101
   : : : : | : | : : : | : | : : | : | : : | : | : : | : | : : |
Db 221 GKTLLAKAV---AGEASVPFFSMGSSFIEMFVLGASRVRLDFIAKKEAPSIIFDEI 277

Qy 102 DQGLSLGSERAHADFDSHLDD---ALNRSLAEQSSAG 135
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 278 D---AIGKSRAGGMISGNDEREOTLNOLLAEMDGF 311

```

RESULT 10  
H69789  
probable alcohol dehydrogenase (EC 1.1.1.-) ydJL [similarity] - Bacillus subtilis  
N:Alternate names: l'-iditol 2'-dehydrogenase homolog ydJL; sorbitol dehydrogenase homolog  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Sep-2000  
C:Accession: H69789

A: Cross-references: GB:AE001471; GB:AE001439; NID:g4154880; PIDN:AAD05932.1; PID:g4154880  
A: Experimental source: strain J99



```
C:Genetics:
A:Gene: Atcu0453
A:Map position: circular chromosome

Query Match      11.0%   Score 75.5; DB 2; Length 248;
Best Local Similarity 31.2%; pred. No. 3.6;
Matches 30; Conservative 6; Mismatches 41; Indels 19; Gaps:

Qy    51..GVLDAGDCGVR-----VCPVGQTATR--EVHITLVQVGEQALFRVGRKAPLAFLD 99
      ||||| ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db    32..GVLDIAADGSLRTDKDPFISVYTDGSKLIEGLELRSLASPGQLDIVFEAGVTTAHAVTD 91
      ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |

Qy    100 -RTDQGLSGSERAHADFSLHD-----DALNRS 127
       ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db    92..TETDESVLGMPATDTATFEFHLDNALRQTGDALNDS 127

Search completed: July 18, 2002, 14:17:35
Job time: 9545 sec
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:31:53 ; Search time 45.63 Seconds  
(without alignments)  
114.555 Million cell updates/sec

Title: US-09-749-185-5  
Perfect score: 687  
Sequence: 1 MSFLNSEELAFRIPVELRYE.....FDSLHLDLALNRLSIAEFSAG 135

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	90.5	13.2	484	1	HXK1_SCHPO
2	78.5	11.4	638	1	FTSH_HELPF
3	77	11.2	632	1	AGPA_RHIME
4	76.5	11.1	632	1	FTSH_HELPJ
5	75.5	11.0	632	1	FTSH_HELPY
6	73	10.6	702	1	FTSH_MYCPE
7	73	10.6	709	1	FTSH_MYCPN
8	72.5	10.6	478	1	AMP2_RAT
9	71.5	10.4	655	1	YAGF_ECOLI
10	71	10.3	891	1	POL2_BAMWN
11	70.5	10.3	478	1	AMP2_HUMAN
12	70.5	10.3	478	1	AMP2_MOUSE
13	70.5	10.3	536	1	ENTE_ECOLI
14	70.5	10.3	551	1	SNA4_MOUSE
15	69.5	10.1	552	1	SMA4_HUMAN
16	69.5	10.1	552	1	SNA4_PIG
17	69.5	10.1	552	1	SNA4_RAT
18	69	10.0	371	1	RIBD_CHLMU
19	69	10.0	3591	1	FHAB_BORPE
20	68.5	10.0	476	1	UL36_HCMVA
21	68.5	10.0	853	1	CNRR_BOVIN
22	68.5	10.0	1134	1	CVAL_BOVIN
23	68	9.9	307	1	SCRK_KLEPN
24	68	9.9	662	1	YHEL_SCHMA
25	66.5	9.7	305	1	HKCT_ECOLI
26	66.5	9.7	3133	1	HNCT_BOMMO
27	66	9.6	430	1	AROA_STAUA
28	66	9.6	491	1	GABD_RHISN
29	66	9.6	539	1	TCFZ_CAEEL
30	66	9.6	930	1	ITHA_HUMAN
31	65.5	9.5	624	1	GLMS_MYCLE
32	65.5	9.5	1063	1	POLS_RUBVR
33	65	9.5	143	1	YW84_MYCTU

34	65	9.5	310	1	VCAP_BPT3	P20324 bacterioph
35	65	9.5	468	1	STHA_MYCTU	O07212 mycobacteri
36	65	9.5	568	1	PTLB_LACIA	P23531 lactococcus
37	65	9.5	797	1	AF32_HUMAN	O94W6 homo sapien
38	64.5	9.4	289	1	YEC4_EBV	P03235 Epstein-Bar
39	64.5	9.4	540	1	CH60_THEBR	Q60024 thermoanaer
40	64	9.3	73	1	GP60_BPSP1	O48414 bacterioph
41	64	9.3	259	1	RPOD_HALMA	Q00813 haloarcula
42	64	9.3	676	1	YMEH_CAEEL	P54813 caenorhabdi
43	63.5	9.2	137	1	YSCB_YEREN	O01243 versinia en
44	63.5	9.2	367	1	NOLF_RHIME	P25196 rhizobium m
45	63.5	9.2	499	1	FEAB_ECOLI	P80668 escherichia

ALIGNMENTS

RESULT 1  
HXK1\_SCHPO  
ID HXK1\_SCHPO STANDARD; PRT; 484 AA.  
AC Q09756;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Hexokinase 1 (EC 2.7.1.1).  
GN HXK1-OR SPAC24H6.04.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=96140736; PubMed=8549830;  
RA Petit T., Blazquez M.A., Gancedo C.;  
RT "Schizosaccharomyces pombe possesses an unusual and a conventional  
RT hexokinase: biochemical and molecular characterization of both  
RT hexokinases";  
RL FEBS Lett. 378:185-189(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: HAS LOW AFFINITY FOR GLUCOSE AND SOME OF ITS ANALOGS.  
CC -!- CATALYTIC ACTIVITY: ATP + D-hexose = ADP + D-hexose 6-phosphate.  
CC -!- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.  
CC -!- SUBUNIT: MONOMER.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X92894; CAA63487.1; -;  
DR EMBL; 254142; CAA90848.1; -;  
DR HSSP; P19367; LHKC.  
DR InterPro; IPR001312; Hexokinase.  
DR Pfam; PF00349; hexokinase; 1.  
DR PRINTS; PR00475; HEXOKINASE.  
DR PRODOM; PD001109; Hexokinase; 1.  
DR PROSITE; PS00378; HEXOKINASES; 1.  
DR TRANSFERASE; Kinase; Glycolysis; ATP-binding; Multigene family.  
KW TRANSFERASE; Kinase; Glycolysis; ATP (BY SIMILARITY).  
FT BINDING 115 115 ATP (BY SIMILARITY).  
FT DOMAIN 154 180 GLUCOSE-BINDING (POTENTIAL).  
SQ SEQUENCE 484 AA; 53597 MW; 165500F19E6BBB0F CRC64;

Query Match 13.2%; Score 90.5; DB 1; Length 484;  
Best Local Similarity 27.0%; Pred. No. 0.11;

Matches 38; Conservative 19; Mismatches 55; Indels 29; Gaps 7;  
Qy 8 ELAFRIPVELRYETVDPYAVRLTEHL---PGDAPV--TWVFGRELLVEG-----VLDA 55  
Db 35 EEQFTIPPELLHRYTDRVSELYKGLTNPQGDVPMVPTWIGTP---DGNHSGSYLALDL 91  
Qy 56 AGDGVRCVPGVQATREHVT-----LQVSEQALFRVKGAPLAFLDRT---D 102  
Db 92 GGT-NLRVCAVEVGNGKFDITQKYLRLPQELKVTREALFDYIADCIKKEVEEHPGKS 150  
Qy 103 QGLSLGSERAHADFDHLDLDA 123  
Db 151 QNLEIGFTFSYPCVQVRSINDA 171

RESULT 2  
FTSH\_HELPF  
ID FTSH\_HELPF STANDARD; PRT; 638 AA.  
AC O32617;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Cell division protein ftsh homolog (EC 3.4.24.-).  
GN FTSH.  
OS Helicobacter felis.  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OX NCBI\_TaxID=214;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 49179;  
RA Bayle D., Wängler S., Weitzenecker T., Volz J., Steinhilber W.,  
R Przybylski M., Sachs G., Schafer K.P., Melchers K.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE  
(BY SIMILARITY).  
CC -!- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41 (ZINC  
METALLOPROTEASE).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
EMBL; AJ001932; CAA05102.1; -  
DR InterPro: IPR003593; AAA.  
DR InterPro: IPR003960; AAA\_sub.  
DR InterPro: IPR003959; AAA\_subfam.  
DR InterPro: IPR000642; Peptidase\_M41.  
DR Pfam; PF00004; AAA; 1.  
DR SMART; SM00382; Peptidase\_M41; 1.  
DR PROSITE; PS00674; AAA; 1.  
KW Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;  
Zinc.  
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 16 36 POTENTIAL.  
FT DOMAIN 37 122 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 123 143 POTENTIAL.  
FT DOMAIN 144 638 CYTOPLASMIC (POTENTIAL).  
FT NP\_BIND 216 223 ATP (POTENTIAL).  
FT METAL 440 440 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 441 441 BY SIMILARITY.  
FT METAL 444 444 ZINC (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 638 AA; 70245 MW; CAA4818021A24A0 CRC64;

Query Match 11.4%; Score 78.5; DB 1; Length 638;  
Best Local Similarity 29.9%; Pred. No. 2.4;  
Matches 29; Conservative 16; Mismatches 41; Indels 11; Gaps 4;  
Qy 44 GRELLVEGLDAAGDGVRCVPGVQATREHVTITLQVSEQALFRVKG--APLAFLDRT 101  
Db 221 GKTLLAKAV---AGEASVPFFSMGSSFIEMFVGLGASRVRLDFDIKKEAPSIIFIDEI 277  
Qy 102 DOGLSLGSERAHADFDHLDLDD---ALNRSLAEQESAG 135  
Db 278 D---AIGKSRAGGMISGNDEREQTLNOLLAEQMDGFG 311

RESULT 3  
AGPA\_RHIME  
ID AGPA\_RHIME STANDARD; PRT; 693 AA.  
AC Q9X4Y1;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Periplasmic alpha-galactoside binding protein precursor.  
GN AGPA OR RB1567 OR SM21647.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OG Plasmid pSymB (megaplasmid 2).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RA Gage D.J., Long S.R.;  
RT "Alpha-galactoside uptake in Rhizobium meliloti: isolation and  
characterization of agpa, a gene encoding a periplasmic binding  
protein required for melibiose and raffinose utilization.";  
RL J. Bacteriol. 180:5739-5748(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396508; PubMed=11481431;  
RA Finan T.M., Weidner S., Hernandez-Lucas I., Becker A., Gouzy J.,  
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Gouzy J.,  
RA Golding B., Puehler A.;  
RT "The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-  
fixing endosymbiont Sinorhizobium meliloti".  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
CC -!- FUNCTION: PROBABLY INVOLVED IN AN ALPHA-GALACTOSIDES INTAKE  
TRANSPORT SYSTEM.  
CC -!- SUBCELLULAR LOCATION: Periplasmic.  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
PROTEIN FAMILY 5.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
EMBL; AF119834; AAD26274.1; -  
DR EMBL; AL603647; CAC49966.1; -  
DR InterPro: IPR000914; SBP\_bac\_5.  
DR Pfam; PF00496; SBP\_bac\_5; 2.  
DR PROSITE; PS01040; SBP\_BACTERIAL\_5; FALSE\_NEG.  
KW Sugar transport; Transport; Periplasmic; Signal; Plasmid;  
Complete proteome.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 693 PERIPLASMIC ALPHA-GALACTOSIDE BINDING  
PROTEIN.  
FT CONFLICT 397 397 E -> K (IN REF. 1).  
SQ SEQUENCE 693 AA; 77407 MW; 76F9B95708C2DF9F CRC64;

```
Query Match 11.2%; Score 77; DB 1; Length 693;
Best Local Similarity 29.5%; Pred. No. 3.8;
Matches 33; Conservative 18; Mismatches 39; Indels 22; Gaps 6;

QY 30 TFLPLGDPATVWVGRRELLVE--GVLDAGDGDVVRVCPVGTATREVHITLVQSGEALF 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 453 TIYYPHD-----LEGAKVLKGVKLTGDTGNGFVNF-PAGKLGKRGDVEIVLLVNSDYSTD 506
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 88 R-----VG---KAPLLAFLDRTDGLSLGSERAHADFDHLDLALNSLAE 130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 507 RNLAEGVGVQMEKGLRLVNLALD-----GKQDAANYAGRFDMWIRHTAE 553
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
FTSH_HELPJ
ID FTSH_HELPJ STANDARD; PRT; 632 AA.
AC Q92M66;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division protein ftsh homolog (EC 3.4.24.-).
GN FTSH OR JHP0356.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE
CC (BY SIMILARITY).
CC -1- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001471; AAD05932.1; -
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003960; AAA.sub.
DR InterPro; IPR003959; AAA.subfam.
DR InterPro; IPR000642; Peptidase_M41.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF01434; Peptidase_M41; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00674; AAA; 1.
KW Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;
KW Zinc; Complete proteome.
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 10 31 POTENTIAL.
FT DOMAIN 32 116 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 117 137 POTENTIAL.
FT DOMAIN 138 632 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 210 217 ATP (POTENTIAL).
FT METAL 434 434 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 435 435 BY SIMILARITY.
```

```
FT METAL 438 438 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 632 AA; 69765 MW; 2FB67B43C51559FB CRC64;

Query Match 11.1%; Score 76.5; DB 1; Length 632;
Best Local Similarity 29.9%; Pred. No. 3.8;
Matches 29; Conservative 15; Mismatches 42; Indels 11; Gaps 4;

QY 44 GRELLVEGLDAAGDGVVRVCPVGTATREVHITLVQSGEALFRVKG--APLLAFLDRT 101
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 215 GKTLAKAV---AGEAHVPFFSMGSSFIENFVGLGSRVDRDLFTAKKQAPSIIIDEI 271
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 102 DQGLSLGSERAHADFDHLDL---ALNRLAEQESAG 135
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 272 D---AIGKRAAGMGWISGNDEREQTLNQLLAEMDGF 305
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
FTSH_HELPJ
ID FTSH_HELPJ STANDARD; PRT; 632 AA.
AC P71408; O07679; O48268;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division protein ftsh homolog (EC 3.4.24.-).
GN FTSH-OR HP1069.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
SEQUENCE FROM N.A.
STRAIN=NCTC 11639 / UA802;
MEDLINE=97047972; PubMed=8892813;
RA Ge Z., Taylor D.E.;
RT "Sequencing, expression, and genetic characterization of the
RT Helicobacter pylori ftsh gene encoding a protein homologous to
RT members of a novel putative ATPase family."
RL J. Bacteriol. 178:6151-6157(1996).
RN [2]
SEQUENCE FROM N.A.
STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=97386403; PubMed=9244252;
RA Beier D., Spohn G., Rappuoli R., Scarlato V.;
RT "Identification and characterization of an operon of Helicobacter
RT pylori that is involved in motility and stress adaptation."
RL J. Bacteriol. 179:4676-4683(1997).
RN [3]
SEQUENCE FROM N.A.
STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watney L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
RN [4]
SEQUENCE OF 324-632 FROM N.A.
STRAIN=69A;
RX MEDLINE=96132941; PubMed=8550601;
RA Melchers K., Weitzenegger T., Buhmann A., Steinhilber W.,
RA Sachs G., Schafer K.P.;
RT "Cloning and membrane topology of a P type ATPase from Helicobacter
RT pylori."
RL J. Biol. Chem. 271:446-457(1996).
CC -1- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE
```



RESULT 8  
AMP2\_RAT  
ID AMP2 RAT  
STANDARD:  
PRT: 478 AA.

OX NCBI\_TaxID=10116;

..... 22

..... 23

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC use by non-profit institutions as long as its content is in no way

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

.....

DR MEROPS; M41.001; -.

```
DR      InterPro; IPR003960; AAA_sub.
DR      InterPro; IPR003960; AAA_sub.
```

DR InterPro; IPR000642; peptidase\_M41.  
PF PF00004

DR Pfam; PF01434; peptidase\_M41; 1.  
DR SMART; SM00202. 1.1.1

DK FROSTIE, FS000/4, AAA, I.  
KW Cell division: ATP-binding: Transmembrane: Hydrophobic: Metalloprotease:

FT	DOMAIN	1	25	CYTOPLASMIC (POTENTIAL)
FT	DOMAIN	1	25	CYTOPLASMIC (POTENTIAL)

ET	DOMAIN	47	171	EXTRACELLULAR (POTENTIAL)
ET	DOMAIN	47	171	EXTRACELLULAR (POTENTIAL)

FT	DOMAIN	193	709	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	193	709	CYTOPLASMIC (POTENTIAL).

FT METAL 490 490 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 494 494 ZINC (CATALYTIC) (BY SIMILARITY).

Query Match  
10.6%; Score 73; DB 1; Length 709;

Matches 29; Conservative 17; Mismatches 32; Indels 16; Gaps 6;

QY 44 GRELLVEGLDAAGDGDVRVCPVGQ-TATREVHITLQVGSEA--LFRVGK--APLLAFL 98

UD 273 GKTLAKAVAGEAG-----VPTFQSTGSGFEDMLVGVGAKRVRDLFNKAKKAPCIIFI 326







```
FT METAL 262 262 COBALT 1 AND 2 (BY SIMILARITY).
FT METAL 331 331 COBALT 1 (BY SIMILARITY).
FT METAL 364 364 COBALT 1 (BY SIMILARITY).
FT METAL 459 459 COBALT 1 AND 2 (BY SIMILARITY).
SQ SEQUENCE 478 AA; 52921 MW; BBB9A2AFC19952E8 CRC64;

Query Match 10.3%; Score 70.5; DB 1; Length 478;
Best Local Similarity 28.8%; Pred. No. 11;
Matches 32; Conservative 19; Mismatches 35; Indels 25; Gaps 8;

QY 17 LRVETDPYAVRTTF--HLPG---DAPVTWVFG--RELLVEGVLDAAAGD-----DVRV 63
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 LQYDDI----CKIDFGTHISGRIIDCAFTVTFNPKYDILLTAVKDATNTGKCGAGIDVRL 297
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 64 CPVCGQTATREHVITLQVSGEALFRVCKAPLLAFLDRTDGLSLGSRANA 114
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 CDVGE-AIQEVMSYEVEIDGKTYQV--KPI-----RNLNHSGISGYPYRIHA 340
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
ENTE_ECOLI
ID ENTE_ECOLI STANDARD; PRT; 536 AA.
AC P10378; P15049; P77773;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2,3-dihydroxybenzoate-AMP ligase (EC 6.3.2.-) (Dihydroxybenzoic acid-
activating enzyme) (Enterobactin synthetase component E) (Enterobactin
synthase E).
DE ENTE OR B0594.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89290355; PubMed=2525505;
RA Staab J.F., Elkins M.F., Earhart C.F.;
RT "Nucleotide sequence of the Escherichia coli entE gene.";
RL FEMS Microbiol. Lett. 50:15-19(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
Fiederspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-8 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90236256; PubMed=2110093;
RA Elkins M.F., Earhart C.F.;
RT "Opacity factor from group A streptococci is an apoprotease.";
RL FEMS Microbiol. Lett. 56:35-40(1988).
RN [5]
RP SEQUENCE OF 393-546 FROM N.A.
RX MEDLINE=89123155; PubMed=2521622;
RA Liu J., Duncan K., Walsh C.T.;
RT "Nucleotide sequence of a cluster of Escherichia coli enterobactin
biosynthesis genes: identification of entA and purification of its
product 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase.";
RL J. Bacteriol. 171:791-798(1989).
```

OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/J; TISSUE=Lung;  
RX MEDLINE=97311184; PubMed=9166592;  
RY Anna C.H., Devereux T.R.;  
RT "Sequence and chromosomal mapping of the mouse homolog (Madh4) of the  
human DPC4/MADH4 gene."  
RL Mamm. Genome 8:443-444(1997).  
RN [2]  
RP SEQUENCE OF 67-551 FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Lung;  
RX MEDLINE=21085660; PubMed=11217851;  
RY Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
RA Saito T., Okazaki Y., Gojori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
RN [3]  
RP REVIEW.  
RX MEDLINE=20175825; PubMed=10708952;  
RY Weinstein M., Yang X., Deng C.-X.;  
RT "Functions of mammalian Smad genes as revealed by targeted gene  
disruption in mice."  
RL Cytokine Growth Factor Rev. 11:49-58(2000).  
CC -1- FUNCTION: COMMON MEDIATOR OF SIGNAL TRANSDUCTION BY TGF-BETA  
SMAD (CO-SMAD) (BY SIMILARITY). SUPERFAMILY: SMAD4 IS THE COMMON  
AS A TUMOR SUPPRESSOR.  
CC -1- SUBUNIT: MAY FORM TRIMERS WITH RECEPTOR-REGULATED SMAD (R-SMAD).  
CC INTERACTS WITH C-SKI, MSK1 AND ATEP2 (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: IN THE CYTOPLASM IN THE ABSENCE OF LIGAND;  
MIGRATION TO THE NUCLEUS WHEN COMPLEXED WITH R-SMAD (BY  
SIMILARITY).  
CC -1- TISSUE SPECIFICITY: UBICUITOUS.  
CC -1- SIMILARITY: BELONGS TO THE DWARFIN/SMAD FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 DWA/WH1 DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 DMB/WH2 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U79748; AAB57905.1; -;  
DR EMBL: AK004804; BAB23576.1; -;  
DR HSSP: Q13485; IDDI.  
DR MGD: MGI:894293; Madh4.  
DR InterPro: IPR001132; Dwarf1n.  
DR InterPro: IPR003619; Dwarf1n\_A.  
DR Pfam: PF00968; Dwarf1n; 1.  
DR SMART: SM00523; DWA; 1.  
DR SMART: SM00524; DWB; 1.  
KW Transcription regulation; Multigene family.  
RN DOMAIN 31 140 DWA.

FT DOMAIN 274 319 SAD.  
FT DOMAIN 320 529 DWB.  
FT DOMAIN 450 465 POLY-ALA.  
FT CONFLICT 257 257 S -> A (IN REF. 2).  
FT CONFLICT 292 292 R -> P (IN REF. 2).  
SQ SEQUENCE 551 AA; 60417 MW; 0835EF8BD9C1C980 CRC64;  
Query Match 10.3%; Score 70.5; DB 1; Length 551;  
Best Local Similarity 32.0%; Pred. No. 13;  
Matches 31; Conservative 14; Mismatches 27; Indels 25; Gaps 8;  
QY 5 VSEELAFRIPV-----ELRYETDPTVAVRLTFHLPGDAPVTWVFGRELLVEGVLD 54  
DB 303 VHNELAQPPISNHPAPEWCSIAFYFEND-VQVGETFKVPSSCPV-----VTVDGYVD 354  
QY 55 AAGDGDVRCVPGQTATREHITLVGSEQALFRVGK 91  
DB 355 PSG-GD-RFC-LGQLS--NVHRTAI--ERARLHIGK 384  
RESULT 15  
SMA4\_HUMAN  
ID SMA4\_HUMAN STANDARD; PRT; 552 AA.  
AC Q13485;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE Mothers against decapentaplegic homolog 4 (SMAD 4) (Mothers against  
DE DPP homolog 4) (Deletion target in pancreatic carcinoma 4) (HSMAD4).  
GN MAD4 OR SMAD4 OR DPC4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANT CARCINOMA HIS-493.  
RC TISSUE=Fetal brain;  
RX MEDLINE=9614684; PubMed=8553070;  
RA Hahn S.A., Schutte M., Shamsul Hoque A.T.M., Moskaluk C.A.,  
da Costa L.T., Rozenblum E., Weinstein C.L., Fischer A., Yeo C.J.,  
RA Hruban R.H., Kern S.E.;  
RT "DPC4, a candidate tumor suppressor gene at human chromosome  
RT 18q21.1".  
RL Science 271:350-353(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=96371046; PubMed=8774881;  
RA Zhang Y., Feng X.-H., Wu R.-Y., Derynck R.;  
RT "Receptor-associated Mad homologues synergize as effectors of the TGF-  
RT beta response".  
RL Nature 383:168-172(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97253203; PubMed=9098646;  
RA Moskaluk C.A., Hruban R.H., Schutte M., Lietman A.S., Smyrk T.,  
RA Fusaro L., Fusaro R., Lynch J., Yeo C.J., Jackson C.E., Lynch H.T.,  
RA Kern S.E.;  
RT "Genomic sequencing of DPC4 in the analysis of familial pancreatic  
RT carcinoma".  
RL Diagn. Mol. Pathol. 6:85-90(1997).  
RN [4]  
RP FUNCTION.  
RX MEDLINE=98051189; PubMed=9389648;  
RA Liu F., Pouppont C., Massague J.;  
RT "Dual role of the Smad4/DPC4 tumor suppressor in TGFbeta-inducible  
RT transcriptional complexes".  
RL Genes Dev. 11:3157-3167(1997).  
RN [5]  
RP CHARACTERIZATION OF SAD DOMAIN.  
RX MEDLINE=20102728; PubMed=10636916;  
RA de Caestecker M.P., Yahata T., Wang D., Parks W.T., Huang S.,

RA Hill C.S., Shioda T., Roberts A.B., Lechleider R.J.;  
RT "The Smad4 activation domain (SAD) is a proline-rich, p300-dependent  
RL transcriptional activation domain.";  
RN J. Biol. Chem. 275:2115-2122(2000).  
RP [6]  
RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 319-543.  
RA MEDLINE-97357157; PubMed-9214508;  
R Shi Y., Hata A., Lo R.S., Massague J., Pavletich N.P.;  
RT "A structural basis for mutational inactivation of the tumour  
RL suppressor Smad4.";  
RN Nature 388:87-93(1997).  
RP [7]  
RX X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 285-552.  
RA MEDLINE-20113477; PubMed-10647180;  
R Qin B., Lam S.S., Lin K.;  
RT "Crystal structure of a transcriptionally active Smad4 fragment.";  
RL Structure 7:1493-1503(1999).  
RN [8]  
RX X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS) OF 273-552.  
RA MEDLINE-21127490; PubMed-11224571;  
R Chacko B.M., Qin B., Correia J.J., Lam S.S., de Caestecker M.P.,  
RT "The L3 loop and C-terminal phosphorylation jointly define Smad  
RL protein trimerization.";  
RN Nat. Struct. Biol. 8:248-253(2001).  
RP [9]  
RX REVIEW.  
RA MEDLINE-98431667; PubMed-9759503;  
R Massague J.;  
RT "TGF-beta signal transduction.";  
RL Annu. Rev. Biochem. 67:753-791(1998).  
RN [10]  
RX REVIEW.  
RA MEDLINE-20112302; PubMed-10647776;  
R Verschuuren K., Huybrecock D.;  
RT "Remarkable versatility of Smad proteins in the nucleus of  
RL transforming growth factor-beta activated cells.";  
RN Cytokine Growth Factor Rev. 10:187-199(1999).  
RP [11]  
RX REVIEW.  
RA MEDLINE-20175821; PubMed-10708948;  
R Wana J.F., Attisano L.;  
RT "The Smad pathway.";  
RL Cytokine Growth Factor Rev. 11:5-13(2000).  
RN [12]  
RX REVIEW.  
RA MEDLINE-20175822; PubMed-10708949;  
R Miyazono K.;  
RT "TGF-beta signaling by Smad proteins.";  
RL Cytokine Growth Factor Rev. 11:15-22(2000).  
CC -!- FUNCTION: COMMON MEDIATOR OF SIGNAL TRANSDUCTION BY TGF-BETA  
CC (TRANSFORMING GROWTH FACTOR) SUPERFAMILY; SMAD4 IS THE COMMON  
CC SMAD (CO-SMAD). PROMOTES BINDING OF THE SMAD2/SMAD4/FAST-1 COMPLEX  
CC TO DNA AND PROVIDES AN ACTIVATION FUNCTION REQUIRED FOR SMAD1 OR  
CC SMAD2 TO STIMULATE TRANSCRIPTION. MAY ACT AS A TUMOR SUPPRESSOR.  
CC -!- SUBUNIT: MAY FORM TRIMERS WITH RECEPTOR-REGULATED SMAD (R-SMAD).  
CC INTERACTS WITH C-SKI, MSI1 AND ATE2.  
CC -!- SUBCELLULAR LOCATION: IN THE CYTOPLASM IN THE ABSENCE OF LIGAND;  
CC MIGRATION TO THE NUCLEUS WHEN COMPLEXED WITH R-SMAD.  
CC -!- DISEASE: DEFECTS IN SMAD4 ARE A CAUSE OF PANCREATIC CARCINOMA.  
CC -!- SIMILARITY: BELONGS TO THE DWARFIN/SMAD FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 DWB/MH1 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 DWB/MH2 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF045447; AAC03051.1; -.

DR EMBL; AF045438; AAC03051.1; JOINED.  
DR EMBL; AF045439; AAC03051.1; JOINED.  
DR EMBL; AF045440; AAC03051.1; JOINED.  
DR EMBL; AF045441; AAC03051.1; JOINED.  
DR EMBL; AF045442; AAC03051.1; JOINED.  
DR EMBL; AF045443; AAC03051.1; JOINED.  
DR EMBL; AF045444; AAC03051.1; JOINED.  
DR EMBL; AF045445; AAC03051.1; JOINED.  
DR EMBL; AF045446; AAC03051.1; JOINED.  
DR EMBL; U44378; AAA91041.1; -.  
DR PDB; 1YGS; 08-NOV-98.  
DR PDB; 1DDI; 24-NOV-99.  
DR PDB; 1G88; 29-NOV-00.  
DR MIM; 600993; -.  
DR InterPro; IPR001132; DwarfIn.  
DR InterPro; IPR003619; DwarfIn\_A.  
DR Pfam; PF00968; DwarfIn; 1.  
DR SMART; SM00523; DWA; 1.  
DR SMART; SM00524; DWB; 1.  
KW Transcription regulation; Phosphorylation; Multigene family;  
KW Disease mutation; 3D-structure.  
FT DOMAIN 31 140 DWA.  
FT DOMAIN 275 320 SAD.  
FT DOMAIN 321 530 DWB.  
FT DOMAIN 451 466 POLY-ALA.  
FT VARIANT 493 493 D -> H (IN PANCREATIC CARCINOMA).  
FT /FTID=VAR\_011380.  
SQ SEQUENCE 552 AA; 60439 MW; 7EE3C4647712DA90 CRC64;  
  
Query Match 10.1%; Score 69.5; DB 1; Length 552;  
Best Local Similarity 30.9%; Pred. No. 16;  
Matches 30; Conservative 15; Mismatches 27; Indels 25; Gaps 8;  
  
QY 5 VSEELAFRIPV-----ELRYETVDPYAVRLTFLPGDAPVTWVFGRELLVEGVLD 54  
DB 304 VHNELAFQPPISNHPAPEYWCSTAYFEMD-VQVGETFKVPSSCPI-----VTVDGYVD 355  
  
QY 55 AAGDGDVRCVPGQTATREHVITLQVGEALFRVGK 91  
DB 356 PSG-GD-RFC-IGQLS--NVHRTAI--ERARLHIGK 385  
  
Search completed: July 18, 2002, 14:31:54  
Job time: 943 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: July 18, 2002, 14:30:54 ; Search time 140.15 Seconds  
(without alignments)  
166.638 Million cell updates/sec  
Title: US-09-749-185-5  
Perfect score: 687  
Sequence: 1 MSFLVSEELAFRIPVELRYE.....FDSLDDALNRLSLAEQSQSAG 135  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*  
1: sp-archaea:\*  
2: sp-bacteria:\*  
3: sp-fungi:\*  
4: sp-human:\*  
5: sp-invertebrate:\*  
6: sp-mammal:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*  
15: sp-rvirois:\*  
16: sp-bacteriap:\*  
17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query Match %	Length	DB ID	Description	Query Match	Best Local Similarity	Mismatches	Indels
1	687	100.0	135	2	Q9F9B6 streptomyc	100.0%	Score 687;	DB 2;	Length 135;
2	498	72.5	136	2	Q9X902 streptomyc	100.0%	Pred. No. 1.2e-60;		
3	481	70.0	135	2	Q9F9B7 streptomyc	135;	Conservative	0;	Mismatches
4	464	67.5	136	2	P95753 streptomyc	0;		0;	Gaps
5	446	64.9	145	2	Q9F9B5 streptomyc	1	MSFLVSEELAFRIPVELRYEVDPAVRLTTHLPDAPVTWVFGRELVEGVLDAGDGD 60		
6	223	32.5	159	2	Q9L268 streptomyc	1	MSFLVSEELAFRIPVELRYEVDPAVRLTTHLPDAPVTWVFGRELVEGVLDAGDGD 60		
7	203	29.5	142	2	Q9S2F7 streptomyc	61	VRVCPVGQTATREHVITLVQVSEQALFRVKGAPLAFLDRTDQGLSLGSERAHADFDSHL 120		
8	155	22.6	138	2	Q9X7M8 streptomyc	61	VRVCPVGQTATREHVITLVQVSEQALFRVKGAPLAFLDRTDQGLSLGSERAHADFDSHL 120		
9	134	19.5	156	2	Q9FC07 streptomyc				
10	117	17.0	142	2	Q9X7R1 streptomyc				
11	107.5	15.6	126	2	Q9RKC9 streptomyc				
12	82	11.9	741	16	Q9H2M0 pseudomonas				
13	79	11.5	862	2	Q9X8M6 streptomyc				
14	78	11.4	258	10	Q9SS82 arabidopsis				
15	78	11.4	346	16	Q34788 bacillus su				
16	77.5	11.3	459	16	Q98DN9 rhizobium l				

17	76.5	11.1	291	16	P71922 mycobacteri
18	76.5	11.1	4349	4	Q9NYQ8 homo sapien
19	76	11.1	232	2	O69245 bradyrhizob
20	76	11.1	736	5	Q9W1Y0 drosophila
21	75.5	11.0	2147	2	Q9L950 pseudomonas
22	75	10.9	563	5	Q9NFK2 boophilus m
23	74.5	10.8	344	17	Q9V2C0 pyrococcus
24	74.5	10.8	394	2	Q9LCH7 streptomyc
25	74.5	10.8	571	10	Q43060 porphyridiu
26	74.5	10.8	571	10	Q43061 porphyridiu
27	74.5	10.8	645	16	Q9PNH8 campylobact
28	73.5	10.7	526	16	Q92TW5 rhizobium l
29	73.5	10.7	639	16	Q98C85 rhizobium l
30	73.5	10.7	4155	5	Q9N3R9 caenorhabdi
31	73	10.6	179	2	O9XA56 streptomyc
32	73	10.6	517	16	Q9I0H3 pseudomonas
33	72.5	10.6	509	16	Q9A2N4 caulobacter
34	72.5	10.6	1327	2	Q9RKJ4 streptomyc
35	72	10.5	146	16	Q98JX5 rhizobium l
36	72	10.5	430	4	Q9H1Q0 homo sapien
37	72	10.5	517	4	Q9Y2Q2 homo sapien
38	72	10.5	715	11	O88967 mus musculu
39	72	10.5	716	4	Q9UMR9 homo sapien
40	72	10.5	740	4	Q96I63 homo sapien
41	72	10.5	773	4	Q96TA2 homo sapien
42	71.5	10.4	1241	4	O60500 homo sapien
43	71	10.3	322	16	Q9KEB8 bacillus ha
44	71	10.3	381	2	Q9KXV7 streptomyc
45	71	10.3	389	2	Q9ZNG9 pseudomonas

ALIGNMENTS

RESULT 1  
Q9F9B6 PRELIMINARY; PRT; 135 AA.  
ID O9F9B6  
DC O9F9B6;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE SSGA.  
GN SSGA.  
OS Streptomyces albus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA van Wezel G.P., Rousseau C., Kraal B.;  
RT "Cloning and sequencing of the Streptomyces albus ssgA gene.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF195771; AAG28482.1; -;  
SQ SEQUENCE 135 AA; 14735 MW; 0FCBFA8DB2BA201B CRC64;

Query Match	100.0%	Score 687;	DB 2;	Length 135;
Best Local Similarity	100.0%	Pred. No. 1.2e-60;		
Matches 135;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
Qy	1	MSFLVSEELAFRIPVELRYEVDPAVRLTTHLPDAPVTWVFGRELVEGVLDAGDGD 60		
Db	1	MSFLVSEELAFRIPVELRYEVDPAVRLTTHLPDAPVTWVFGRELVEGVLDAGDGD 60		
Qy	61	VRVCPVGQTATREHVITLVQVSEQALFRVKGAPLAFLDRTDQGLSLGSERAHADFDSHL 120		
Db	61	VRVCPVGQTATREHVITLVQVSEQALFRVKGAPLAFLDRTDQGLSLGSERAHADFDSHL 120		
Qy	121	DDALNRLSLAEQSQSAG 135		
Db	121	DDALNRLSLAEQSQSAG 135		

Mon Jul 22 08:06:43 2002

```
RESULT 2
Q9X902 ID Q9X902 PRELIMINARY; PRT; 136 AA.
AC Q9X902; 1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE REGULATOR.
GN SCQ11.09.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K., Harris D.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RT Kinashi H., Hopwood D.A.; and a detailed genetic and physical map for
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL096823; CAB46964.1; -.
SQ SEQUENCE 136 AA; 14970 MW; 4B67C1F1E0BECC88 CRC64;

Query Match 72.5%; Score 498; DB 2; Length 136;
Best Local Similarity 75.6%; Pred. No. 6.9e-42;
Matches 102; Conservative 9; Mismatches 24; Indels 0; Gaps 0;

QY 1 MSFLVSEELAFRIPVELRYETVDPYAVRLTFHLPDGPVTVWFGRELLVEGVLDAGDGD 60
DB 1 MSFLVSEELAFRIPVELRYETVDPYAVRLTFHLPDGPVTVWFGRELLVEGVLDAGDGD 61
QY 61 VRVCPVGTATREVHITLQVSGEQALFRVKGAPLAFLDRTDQGLSGSERAHADFDSL 120
DB 62 VRIAPVEPEPLAEVLRLQVSGDQALFRSSAAPLVAFLDRTDKLVPFGQEGALADFDSL 121
QY 121 DDALNRLAEQESAG 135
DB 122 DEALDRILAEQESAG 136

Query Match 70.0%; Score 481; DB 2; Length 135;
Best Local Similarity 71.1%; Pred. No. 3.3e-40;
Matches 96; Conservative 14; Mismatches 25; Indels 0; Gaps 0;

QY 1 MSFLVSEELAFRIPVELRYETVDPYAVRLTFHLPDGPVTVWFGRELLVEGVLDAGDGD 60
DB 1 MSFLVSEELAFRIPVELRYETVDPYAVRLTFHLPDGPVTVWFGRELLVEGVLDAGDGD 60
QY 61 VRVCPVGTATREVHITLQVSGEQALFRVKGAPLAFLDRTDQGLSGSERAHADFDSL 120
DB 61 VHIAPADPETFGEVLRLQVSGDQAMFRVGTAPLVAFLDRTDKIVPLGQERSLADFALL 120
QY 121 DDALNRLAEQESAG 135
DB 121 DEALDRILAEQESAG 135

RESULT 4
P95753 ID P95753 PRELIMINARY; PRT; 136 AA.
AC P95753;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SSGA.
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B2682;
RA Shinichi K., Ensign J.;
RT "Cloning and characterization of a gene involved in sporulation and
RT cell division of Streptomyces griseus.";
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B2682;
RA Kawamoto S., Ensign J.C.;
RT "Isolation of mutants of Streptomyces griseus that sporulate in
RT nutrient rich media.";
RL Nippon Hosenkin Gakkaishi 9:124-135(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=B2682;
RA Kawamoto S., Ensign J.C.;
RT "Cloning and characterization of a gene involved in regulation of
RT sporulation and cell division of Streptomyces griseus.";
RL Nippon Hosenkin Gakkaishi 9:136-151(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B2682;
RX MEDLINE=97286526; PubMed=9141673;
RA Kawamoto S., Watanabe H., Hesketh A., Ensign J.C., Ochi K.;
RT "Expression analysis of a ssgA gene product associated with
RT sporulation and cell division in Streptomyces griseus.";
RL Microbiology 143:1077-1086(1997).
DR EMBL; D50051; BAA21558.1; -.
SQ SEQUENCE 136 AA; 14783 MW; C6A28A7823AD7C8B CRC64;

Query Match 67.5%; Score 464; DB 2; Length 136;
Best Local Similarity 65.9%; Pred. No. 1.6e-38;
Matches 89; Conservative 18; Mismatches 28; Indels 0; Gaps 0;

QY 1 MSFLVSEELAFRIPVELRYETVDPYAVRLTFHLPDGPVTVWFGRELLVEGVLDAGDGD 60
DB 2 MSFLVSEELAFRIPVELRYETVDPYAVRLTFHLPDGPVTVWFGRELLVEGVLDAGDGD 61
QY 61 VRVCPVGTATREVHITLQVSGEQALFRVKGAPLAFLDRTDQGLSGSERAHADFDSL 120
```







[illegible]

```
Best Local Similarity 31.7%; Pred. No. 23;
Matches 39; Conservative 6; Mismatches 40; Indels 38; Gaps 6;

QY 17 LRVEYVDPYAVR-----LTFHLPDGPVTVWFGREL-----LVGGVLDAAGDG-DVRV 63
Db 646 LRYAAEVAPTVGRVEDGLATHVP-----PTFEAAREVARAGGDLVLGLLPGPEDGID---- 699
QY 64 CPVGQTATREHVHTTQVSGEQALFRVGVKAPLLAFLDRDTQGLSLGSGERAHADFDSDLDDA 123
Db 700 ----DSTAARVHTAVSGSEAL-----QFLSRPREHTGADVQEFPLATS 740
QY 124 LNR 126
Db 741 WNR 743

RESULT 14
Q9SS82 PRELIMINARY; PRT; 258 AA.
AC Q9SS82;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MZB10.8 PROTEIN.
GN MZB10.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN RP
SEQUENCE FROM N.A.
RC STRAIN=CV, COLJMBTA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.F., Otterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III p1 MZB10 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC009326; AAD56321.1; -.
SQ SEQUENCE 258 AA; 29249 MW; B347996F2F2C07F0 CRC64;

Query Match 11.4%; Score 78; DB 10; Length 258;
Best Local Similarity 29.7%; Pred. No. 6.9;
Matches 30; Conservative 14; Mismatches 35; Indels 22; Gaps 4;

QY 50 EGVLDRAAGDGVRCVPGVQTATREHVHTLQVSGEQALFRVGVKAPLLAFLDRDTQGLSLGS 109
Db 40 DGTSDSDSDPPPK-PEGDTFRQELLARI-----AMIQTSKVRLTDFLDERSEYLTKEA 92
QY 110 ERAHADFD-----SHLDALNRSL-----AEEQSAG 135
Db 93 EEAFAEDKVGEDAMKDLDEASTRILENTESKMQAFESAG 133

RESULT 15
O34788 PRELIMINARY; PRT; 346 AA.
AC O34788;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DEHYDROGENASE.
GN YDJL
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
[1]
RN RP
SEQUENCE FROM N.A.
RC STRAIN=MARBURG 168;
RA Kasahara Y., Nakai S., Ogasawara N., Yata K., Sadaie Y.;
RL DNA Res. 0:0(1997).
```

```

RN RP
SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=980404033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivotla C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter K., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RL Nature 390:249-256(1997).
[3]
RN RP
SEQUENCE FROM N.A.
RC STRAIN=168;
RX Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -I- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
CC EMBL: AB007638; BAA22767.1; -.
DR EMBL: Z99107; CAB12443.1; -.
DR HSSP: P07846; 1SDG.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Complete proteome; Oxidoreductase; Zinc.
SQ SEQUENCE 346 AA; 37341 MW; 5E327192D678F8A2 CRC64;
```

```
Query Match 11.4%; Score 78; DB 16; Length 346;
Best Local Similarity 39.3%; Pred. No. 9.8;
Matches 33; Conservative 6; Mismatches 29; Indels 16; Gaps 5;
```

```
QY 5 VSEELAFRIPVELRYE---TVDP-----YAVRLTFLPGDAPVTVWFG---RELLYEGVL 53
Db 131 VDEELFLKLPDELVSBOGALVPSVAVALYAVRSSKUKAGDKAA--VFGCGPGLLVIEAL 188
QY 54 DAAGDGD---VRVCPVQGTATREV 74
Db 189 KAAGATDIYAVELSPERQOKAEEL 212
```

Search completed: July 18, 2002, 14:30:55  
Job time: 944 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:22:05 ; Search time 4085.76 Seconds  
(without alignments)  
2084.582 Million cell updates/sec

Title: US-09-749-185-6  
Perfect score: 407  
Sequence: 1 atgagcttcctcgtctcgga.....gaggagcagaacgcgcgctg 407

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

GenEmbl:\*

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

#### RESULT: 1

AX007222	LOCUS	AX007222	Sequence 7 from Patent WO00000613.	407 bp	DNA	linear	PAT 06-SEP-2000
AX007222	DEFINITION	Sequence 7 from Patent WO00000613.					
AX007222	ACCESSION	AX007222					
AX007222.1	VERSION	AX007222.1	GI:9995088				
AX007222	KEYWORDS	Streptomyces goldeniensis.					
AX007222	SOURCE	Streptomyces goldeniensis					
AX007222	ORGANISM	Streptomyces goldeniensis					
AX007222	REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.					
AX007222	AUTHORS	1 (bases 1 to 407)					
AX007222	TITLE	Kraal, B., Luiten, R.G. and Van Wezel, G.P.					
AX007222	JOURNAL	Reducing branching and enhancing fragmentation in culturing filamentous microorganisms					
AX007222	JOURNAL	Patent: WO 000613-A 7 06-JAN-2000;					
AX007222	JOURNAL	UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA (NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS VAN (NL)					
AX007222	FEATURES	Location/Qualifiers					
AX007222	source	1. .407					
AX007222	source	/db_xref="taxon:121022"					

1	407	100.0	407	6	AX007222	Sequence
2	407	100.0	408	1	AF195770	Streptomy
3	295	72.5	15441	1	SCQ11	Streptomy
4	259.8	63.8	407	6	AX007218	Sequence
5	259.8	63.8	438	6	AX007216	Sequence
6	259.8	63.8	1513	1	D50051	Streptomyce
7	248.6	61.1	407	6	AX007224	Sequence
8	248.6	61.1	438	1	AF195772	Streptomy
9	243.8	59.9	407	6	AX007220	Sequence
10	243.8	59.9	566	1	AF195771	Streptomy
11	110.4	27.1	35284	1	SCE19A	Streptomy
12	95.6	23.5	38640	1	SCL2	Streptomy
13	94.6	23.2	1410	1	STMLACBG	M28303 S.albus bet
14	78	19.2	32704	1	SC8A11	AL391041 Streptomy
15	59.4	14.6	29625	1	SCH22A	AL159178 Streptomy
16	58.8	14.4	45624	6	AX089419	Sequence
17	58.8	14.4	50000	6	AX089416	Sequence
18	58.2	14.4	44274	1	SCH63	AX089416 Sequence
19	58.2	14.3	110000	2	LMFLCHR32_06	Continuation (7 of
20	57.2	14.1	213050	1	AL646079	AL646079 Ralstonia
21	56.4	13.9	152172	8	AC068924	AC068924 Oryza sat
22	56.2	13.8	110373	2	AP004321	AP004321 Oryza sat
23	56.2	13.8	146017	2	AP003575	AP003575 Oryza sat
24	55.8	13.7	16832	1	AB070943	AB070943 Streptomy
25	55.6	13.7	1258	1	TAU52907	U52907 Thermus aqu
26	55.6	13.7	38543	1	SCBAC16H6	AL596162 Streptomy
27	55.4	13.6	92509	1	AL646086	AL646086 Ralstonia
28	55.2	13.6	37245	1	SC5F2A	AL049587 Streptomy
29	54.8	13.5	23030	1	SC9E12	AL391751 Streptomy
30	54.4	13.4	3231	1	PSEACCB	LI4612 Pseudomonas
31	54.4	13.4	10927	1	AE004898	AE004898 Pseudomon
32	54.2	13.3	11104	1	AE001952	AE001952 Deinococc
33	54	13.3	40901	1	SCC24	AL163003 Streptomy
34	53.8	13.2	45590	1	SCBAC1A6	AL589708 Streptomy
35	53.8	13.2	110000	2	LMFLCHR16_02	Continuation (3 of
36	53.8	13.2	110000	2	LMFLCHR16_03	Continuation (4 of
37	53.8	13.2	110000	2	LMFLCHR36_31	Continuation (32 o
38	53.6	13.2	10029	1	AE005987	AE005987 Caulobact
39	53.4	13.1	32608	1	SCBAC36F5	AL592292 Streptomy
40	53.4	13.1	36583	1	SC5H1	AL049863 Streptomy
41	53.4	13.1	110000	2	LMFLCHR36_07	Continuation (8 of
42	53.2	13.1	34644	1	AF080235	AF080235 Streptomy
43	53.2	13.1	44375	1	SC6G9	AL079356 Streptomy
44	53	13.0	37931	1	SCD10	AL359988 Streptomy
45	53	13.0	140623	8	AC078894	AC078894 Oryza sat

```
gene 1..405
CDS 1..>405
  /gene="ssga"
  /note="strain ATCC of Streptomyces goldeniensis"
  /codon_start=1
  /transl_table=11
  /product="SsgA"
  /protein_id="CAC07387.1"
  /db_xref="GI:995089"
  /translation="MSFLVSELSRIPVELRYETCDPYAVRLTHLPGDAPVTWAFG
  RELIDGGPRPCGDGVHIAPADPTFGEVLIRLQVSDQAMFRVGTAPLVAFLDRDRTD
  KIVPLGQERSLADFALLDEALDRILAEQNAG"
BASE COUNT 56 a 135 c 145 g 71 t
ORIGIN
```

```
Query Match 100.0%; Score 407; DB 6; Length 407;
Best Local Similarity 100.0%; Pred. No. 5.3e-42;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagattctcttcggaagaactctcttcgatttccttcggtgagctgcgttacgag 60
DB 1 ATGAGCTTCTCTCTCGGAAGAACTCTCTCTCCGTATTCCGGTGGAGCTGCGTTACGAG 60
QY 61 acctgtgacctacccgctgcggctgacctttcattctcgcggagatgcccccgtgacc 120
DB 61 ACCTGTGATCCCTACGCCGTGCGGCTGACCTTTTCATCTGCCCGGAGATGCCCGCGTGACC 120
QY 121 tgggcttcgggaggagttgctcatcagcaggaggtccggcgtgcggggacggggac 180
DB 121 TGGCGCTTCGGCGGGAGTGTCTCATCGACGAGGTCCGCGCGCTGCGGGACGGGGAC 180
QY 181 gtccacatcgcccgccgaccggagagcttcggcgaggtcctgacgcctgcagggtg 240
DB 181 GTCCACATCGCCCGCCGACCGAGAGCTTCGGCGAGGTCTGATCGCCCTGCAGGTG 240
QY 241 gggagcagaccagcgatgttcctccgggtcgcaaggcgcgtggtgaccttcctggaccgc 300
DB 241 GGGAGCAGCAGCGGATGTTCGGGTGCGGACGCGCGCGCTGTGTCCTTCTGGACCGC 300
QY 301 acggacaagatctgcgcgtggcgaggagcgttcctcgcgcgaacttcagcgcctgctc 360
DB 301 ACGGACAAGATCTGCGCGTGGGCGAGGAGCTTCCTCTGCCGACTTCGACGCGCTGCTC 360
QY 361 gacgagcctgacgcacatcttcggcagagagcagacagacccgctg 407
DB 361 GACGAGCGCTGACCGCATCTTGGCGGAGGAGCAGAACCGCGGCTG 407
```

```
RESULT 2
AF195770 408 bp DNA linear BCT 01-NOV-2000
LOCUS Streptomyces goldeniensis SsgA (ssgA) gene, complete cds.
DEFINITION AF195770
ACCESSION AF195770.1 GI:11066158
VERSION AF195770.1
KEYWORDS Streptomyces goldeniensis.
SOURCE Streptomyces goldeniensis.
ORGANISM Streptomyces goldeniensis.
REFERENCE Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
  Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
  1 (bases 1 to 408)
AUTHORS van Wezel, G.P., Rousseau, C. and Kraal, B.
TITLE Cloning and sequencing of the Streptomyces goldeniensis ssgA gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 408)
AUTHORS van Wezel, G.P.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Biochemistry, University of Leiden, PO Box
  9502, Leiden 2300 RA, Netherlands
FEATURES
  source
    1..408
    /organism="Streptomyces goldeniensis"
```

```
/db_xref="ATCC:21386"
/db_xref="taxon:121022"
1..408
  /gene="ssga"
  /note="ssga"
  /note="similar to Streptomyces griseus SsgA; possible cell
  division protein"
  /codon_start=1
  /transl_table=11
  /product="SsgA"
  /protein_id="AAG28481.1"
  /db_xref="GI:11066159"
  /translation="MSFLVSELSRIPVELRYETCDPYAVRLTHLPGDAPVTWAFG
  RELIDGGPRPCGDGVHIAPADPTFGEVLIRLQVSDQAMFRVGTAPLVAFLDRDRTD
  KIVPLGQERSLADFALLDEALDRILAEQNAG"
BASE COUNT 57 a 135 c 145 g 71 t
ORIGIN

Query Match 100.0%; Score 407; DB 1; Length 408;
Best Local Similarity 100.0%; Pred. No. 5.3e-42;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgagcttctcttcggaagaactctcttcgatttccttcggtgagctgcgttacgag 60
DB 1 ATGAGCTTCTCTCTCGGAAGAACTCTCTCTCCGTATTCCGGTGGAGCTGCGTTACGAG 60
QY 61 acctgtgacctacccgctgcggctgacctttcattctcgcggagatgcccccgtgacc 120
DB 61 ACCTGTGATCCCTACGCCGTGCGGCTGACCTTTTCATCTGCCCGGAGATGCCCGCGTGACC 120
QY 121 tgggcttcgggaggagttgctcatcagcaggaggtccggcgtgcggggacggggac 180
DB 121 TGGCGCTTCGGCGGGAGTGTCTCATCGACGAGGTCCGCGCGCTGCGGGACGGGGAC 180
QY 181 gtccacatcgcccgccgaccggagagcttcggcgaggtcctgacgcctgcagggtg 240
DB 181 GTCCACATCGCCCGCCGACCGAGAGCTTCGGCGAGGTCTGATCGCCCTGCAGGTG 240
QY 241 gggagcagaccagcgatgttcctccgggtcgcaaggcgcgtggtgaccttcctggaccgc 300
DB 241 GGGAGCAGCAGCGGATGTTCGGGTGCGGACGCGCGCGCTGTGTCCTTCTGGACCGC 300
QY 301 acggacaagatctgcgcgtggcgaggagcgttcctcgcgcgaacttcagcgcctgctc 360
DB 301 ACGGACAAGATCTGCGCGTGGGCGAGGAGCTTCCTCTGCCGACTTCGACGCGCTGCTC 360
QY 361 gacgagcctgacgcacatcttcggcagagagcagacagacccgctg 407
DB 361 GACGAGCGCTGACCGCATCTTGGCGGAGGAGCAGAACCGCGGCTG 407

RESULT: 3
SC011 15441 bp DNA linear BCT 08-JUL-1999
LOCUS Streptomyces coelicolor cosmid Q11.
DEFINITION SC011
ACCESSION AL096823
VERSION AL096823.1 GI:5457267
KEYWORDS abab; bldA regulation; cysA; cystathionine/methionine
  gamma-synthase/lyase; gntR-family; integrated element; korSA;
  lysR-family; membrane protein; phosphotyrosine protein phosphatase;
  pra; pSAM2; ptpA; ssgA; thiamine biosynthesis; thic;
  transcriptional regulator; trasa.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
REFERENCE Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
  Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
  1 (bases 1 to 15441)
AUTHORS Redenbach, M., Kieser, H.M., Denapalite, D., Eichner, A., Cullum, J.,
  Kinashi, H. and Hopwood, D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
  for the 8 Mb Streptomyces coelicolor A3(2) chromosome
```

Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351  
2 (bases 1 to 15441)  
Seeger,K. and Harris,D.  
Unpublished  
3 (bases 1 to 15441)  
James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.  
Direct Submission  
Submitted (08-JUL-1999) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded  
by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL; [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
CDS are numbered using the following system eg SC7B7.01c. SC (S.  
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary  
strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons  
using a specially developed Hidden Markov Model (Krogh et al.,  
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot  
program of Bibb et al., Gene 30:157-66(1984) as implemented at  
<http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the  
correct initiation codon. Where possible we choose an initiation  
codon (atg, gtg, ttg or att) which is preceded by an upstream  
ribosome binding site sequence (optimally 5-13bp before the  
initiation codon). If this cannot be identified we choose the most  
upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions.  
Cosmid Q11 Cosmid Q11 lies between H24 and D78 on the AseI-Q  
genomic restriction fragment.  
Location/Qualifiers  
1..15441  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid Q11"  
complement(1..289)  
/gene="SCQ11.01c"  
complement(<1..289)  
/gene="SCQ11.01c"  
/feature="SCQ11.01c, hypothetical protein, partial CDS, len:  
>95 aa; unknown function, previously sequenced as  
TR:Q53870 (EMBL:U37580), S.coelicolor hypothetical protein  
ORF5 downstream of phosphotyrosine protein phosphatase  
(159 aa). Weak similarity to the N-terminus of Tr:O69888  
(EMBL:AL023797) Streptomyces coelicolor hypothetical  
protein (172 aa), fasta scores: opt: 158 z-score: 206.4  
E(): 0.00036, 46.2% identity in 65 aa overlap. Continues  
as SCH24.40c in cosmid H24 (EMBL:AL049826)"  
/codon\_start=1  
/transl\_table=11  
/label="SCQ11.01c  
/product="hypothetical protein"  
/protein\_id="CAB46956.1"  
/db\_xref="GI:5457268"

translation="MTVPRVVKTRAVLDDGSHLILIKRTKPGVDPYVWTPGGGVEP  
DDTVVADLHREVYELGASIDVVPFCVDTVHIGEDGATGKVKQHFVC"

complement(1..4279)  
/note="sequence corresponding to EMBL:U37580 from 1 to  
4279"  
1..98  
/note="overlap with Streptomyces coelicolor cosmid H24  
(EMBL:AL049826) from 41528 to 41625"  
complement(360..1265)  
/gene="SCQ11.02c"  
/note="abaB"  
complement(360..1265)  
/gene="SCQ11.02c"  
/note="SCQ11.02c, abaB, probable lysR-family  
transcriptional regulator, len: 301 aa; previously  
sequenced as TR:Q53869 (EMBL:U37580) S.coelicolor  
lysR-like protein (301 aa). Highly similar to  
SW:ARAB\_STRAT (EMBL:X91393), arab, Streptomyces  
antibiotic transcriptional regulatory protein (301 aa),  
fasta scores: opt: 1581 z-score: 1813.2 E(): 0, 80.3%  
identity in 300 aa overlap. Similar to many other  
transcriptional regulators. Contains probable  
helix-turn-helix motif at aa 17-38 (Score 1659, +4.84 SD).  
Contains Pfam match to PF00126 HTH\_1, Bacterial regulatory  
helix-turn-helix protein, lysR family and PS00044  
Bacterial regulatory proteins, lysR family signature"  
/codon\_start=1  
/transl\_table=11  
/label="abaB  
/product="putative lysR-family transcriptional regulator"  
/protein\_id="CAB46957.1"  
/db\_xref="GI:5457269"  
/translation="MDLALLFTVTVHRSFTFRAAALLGLSOPANTSQITLRLQRLQ  
RPLFQRQGVPTTIGDELAKHAHPDLALVEIAETGLEDSTLRLHLAGPPEPTA  
ERALPVLGELTGEDQOAFALRASFGNAEETLEAGHGHDLAIGTTPRCALHTATPL  
CDEHVLVATPWAERAGVEDVDTASALKHGVVVEHSLPEVGVASVDFDARA  
SPATVAVDRLAVLACAVAGAGLAVLPRYLCAEALERGDVVALHDPPVPLRYFLV  
RTGLAMPHIAHAEHWLRLRAADWN"  
/gene="SCQ11.02c"  
/note="Match to PF00126 HTH\_1, Bacterial regulatory  
helix-turn-helix protein, lysR family Score 145.96"  
complement(1125..1217)  
/gene="SCQ11.02c"  
/note="PS00044 Bacterial regulatory proteins, lysR family  
signature"  
complement(1518..2696)  
/gene="SCQ11.03c"  
/note="cysA"  
complement(1518..2696)  
/gene="SCQ11.03c"  
/note="SCQ11.03c, cysA, probable cystathionine/methionine  
gamma-synthase/lyase, len: 392 aa; previously sequenced as  
SW:CYSA\_STRCO (EMBL:U37580), cysA, S.coelicolor putative  
cystathionine gamma-lyase (392 aa). Similar to many e.g.  
SW:MEGL\_PSEPU (EMBL:D88554), mdeA, Pseudomonas putida  
methionine gamma-lyase (398 aa), fasta scores: opt: 608  
z-score: 653.6 E(): 4.5e-29, 36.9% identity in 404 aa  
overlap. Highly similar to TR:Q53668 (EMBL:X91393)  
Streptomyces antibiotic hypothetical protein found  
upstream of the abaB gene (232 aa) (87.1% identity in 232  
aa overlap). Contains Pfam match to PF01053  
Cys\_Met\_Met\_PP, Cys/Met metabolism PLP-dependent enzyme  
and PS00868 Cys/Met metabolism enzymes pyridoxal-phosphate  
attachment site"  
/codon\_start=1  
/transl\_table=11  
/label="cysA  
/product="putative cystathionine/methionine  
gamma-synthase/lyase"  
/protein\_id="CAB46958.1"  
/db\_xref="GI:5457270"

translation="MSDSATFDSAGTGERSASAPDGTTRAVRAGLPEPVVKHPTLPF  
PVFAHFLPGDPTGPTYTIDENPTWTRLESAIGLEAFGEAGVETLVPASGMAA15  
SVLFSOLRAGTAVLPDDGYALPLVRAQLAEIVETRTAPTGTDAOLDVLDGAKL11

misc\_feature  
complement(1533..2624)  
/gene="SC011.03c"  
/note="Match to PF01053 Cys\_Met\_Met\_PP, Cys/Met  
metabolism PIP-dependent enzyme Score 268.52"  
complement(2031..2075)  
/gene="SC011.03c"  
/note="PS00868 Cys/Met metabolism enzymes  
pyridoxal-phosphate attachment site"  
complement(2693..3187)  
/gene="SC011.04c"  
/note="ptpA"  
complement(2693..3187)  
/gene="SC011.04c"  
/note="SC011.04c, ptpA, low molecular weight  
protein-tyrosine-phosphatase, len: 164 aa; previously  
sequenced as SW:PTPA\_STRCO (EMBL:U37580), ptpA,  
S.coelicolor low molecular weight  
protein-tyrosine-phosphatase (164 aa). Contains Pfam match  
to PF01451 LMWPC. Low molecular weight phosphotyrosine  
protein phosphatase"  
/codon\_start=1  
/transl\_table=11  
/label=ptpA  
/product="low molecular weight  
protein-tyrosine-phosphatase"  
/protein\_id="CAB46959.1"  
/db\_xref="GI:5457271"  
/translation="MYRVCVTGTCNCRSPMAEAEVRAVEDAGLHLEADSCTG  
GWHEGEGAPRTFAVLADHGYGLDHAARFQOSWFSRLDLVLDVALDAGHLRALRLAPT  
ERDAARVLLRSRDPVAVAGDLDPDPYTGGRDGFEECLEWFEASTGLLAAREQVE  
GRAA"

Query Match 72.5%; Score 295; DB 1; Length 15441;  
Best Local Similarity 82.8%; Pred. No. 1.6e-28;  
Matches 337; Conservative 0; Mismatches 70; Indels 0; Gaps 0;  
QY 1 atgagcttcctcgtcgcggaagaactcctcctcgtattccggtgagctgcgttacgag 60  
Db 5475 ATGAGCTTCTCGTGTCCGAGGAGCTCTTTCGCGCATCCGGTGGAGCTCGCGTACGAG 5534  
QY 61 acctgtatccctacgcgctgcgctgacatttcattcctgcccagagatgcccggtgacc 120  
Db 5535 ACCCGGATCCCTATGCGGTACGCTGACCTTTTCATCTGCCGAGACGCGCGGTGACC 5594  
QY 121 tgggcttcggtcgcggaaggttctcattcgcgaggtccgcggtcgcggtcgcggtgac 180  
Db 5595 TGGGCTTCGCGCGGAGCTGCTGGTACGCGGTGGCGCGCGCTGCGCGGCGGAC 5654  
QY 181 gtccacatcgc 240  
Db 5655 GTGCGCATCGCGCGGTGGAGCGGAGCGGCTGGCGCGGAGTGTGATCCGACTTCAGGTC 5714  
QY 241 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300  
Db 5715 GGCAGCACACGAGCGCTGTCCGTTCTCCGCGCGCGCGCTGGTGGCGCTTCTTCGACCGC 5774  
QY 301 acggacaagatcgtgcgcgtcgc 360  
Db 5775 ACCGACAAGCTGTGCGCTGGGCGAGAGGGCGGCTCGCGCGACTTCGACAGCCACCTC 5834  
QY 361 gacgagcgtcgtgacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 407  
Db 5835 GAGGAGCCCTGGACCGCATCTCTGGCGAGGAGAACAGAGCGCGGCGTG 5881

RESULT 4  
AX007218  
LOCUS AX007218 407 bp DNA linear PAT 06-SEP-2000

DEFINITION  
Accession  
Version  
Keywords  
Source  
Organism

Sequence 3 from Patent WO0000613.  
AX007218  
AX007218.1 GI:9995084

Streptomyces griseus.  
Streptomyces griseus

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 407)  
Kraal, B., Luiten, R.G. and Van Wezel, G.P.  
Reducing branching and enhancing fragmentation in culturing  
filamentous microorganisms  
Patent: WO 0000613-A 3 06-JAN-2000;  
UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA  
(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS  
VAN (NL)

FEATURES  
source  
1..407  
/organism="Streptomyces griseus"  
/db\_xref="taxon:1911"

gene  
1..405  
/gene="ssgA"

CDS  
1..>405  
/gene="ssgA"

/note="strain ATCC of Streptomyces griseus"

/codon\_start=1  
/transl\_table=11  
/product="SsgA"

/protein\_id="CAC07385.1"

/db\_xref="GI:9995085"

/translation="MSFLVSEELSFRIPVELRYEVDGPVYAIRMTFHLPGDAPVTWAFG  
RELLDGLNSPSGDGVHIGTPEGLGDVHRLQVGADRALFRAGTAPLVAFLLDRD  
KLVLPGQHTLGDGDFGNLELDALGRILAEENAG"

BASE COUNT 60 a 139 c 142 g 66 t

ORIGIN

Query Match 63.8%; Score 259.8; DB 6; Length 407;  
Best Local Similarity 77.4%; Pred. No. 1e-23;  
Matches 315; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
QY 1 atgagcttcctcgtcgcggaagaactcctcctcgtattccggtgagctgcgttacgag 60  
Db 1 ATGAGCTTCTCGTGTCCGAGGAGCTCTCGTTCGTTATTCGGTGGAGCTCCGATACGAG 60  
QY 61 acctgtatccctacgcgctgcgctgacatttcattcctgcccagagatgcccggtgacc 120  
Db 61 GTCGGCGATCCGTATGCGATCCCGATACGTTCCACCTTCCCGCGATGCCCTGTGACC 120  
QY 121 tgggcttcggtcgcggaaggttctcattcgcgaggtccgcggtcgcggtcgcggtgac 180  
Db 121 TGGGCTTCGCGCGGAGCTGCTGCGACGGGCTCAACAGCCGCGAGCGGCGGAT 180  
QY 181 gtccacatcgc 240  
Db 181 GTGACATCGCGCGCGAGCGGCGGCTCGGAGATGTCCACATTCGAGTCCAGGTC 240  
QY 241 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300  
Db 241 GGGCGGACCGCTGCGCTGTTCGCGCGGCGGAGCGGACCGCTGGTGGCGTTCCTCGACCGG 300  
QY 301 acggacaagatcgtgcgcgtcgc 360  
Db 301 ACGACAAGCTGTGCGCTGCGCGCGCTCGCGCGAGGACACACGCTGGGTGACTTCGACGGAACCTG 360  
QY 361 gacgagcgtcgtgacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 407  
Db 361 GAGGAGCCCTGGACCGCATCTCTGGCGAGGAGAACAGAGCGCGGCGTG 407

RESULT 5  
AX007216  
LOCUS AX007216 438 bp DNA linear PAT 06-SEP-2000



Qy	181:	gtccacatcgcgcccgcgaccggagacgttccggcgagtctctgatccgcctgcaggtg	240
Db	181	GTCCACATCGCCCCGACCGACCAGGGCGCTGTGCGSACGTCCTCCATCCGGCTCCAGGTG	240
Qy	241	gggagcgacacaggcgaatttcctggggtcgcacggcgccgctgtagccttctctggaccgc	300
Db	241:	GGCGCGAACCAGCGGCCCTTTCCGTGCAGSGCGCCCGCGCTGTGCGCTTCTCCTCGACCGC	300
Qy	301	acgcacaagaatcgtgcgcctggggcaggagcgtttccctgccgcactctcaqcgccctgctc	360
Db	301:	ACGGAACAAGTCGTGCCGCTCGTGTCAGGAAACAGACTCTGGGTGACTTCGAGGACAGCCCTG	360
Qy	361	gacgagcgctggaccgcataccttggccggcaggagcagaacccggctg	407
Db	361	GAGCGCCGCTCGCAAGATCTCTCGCCAGGAGCAGAACCSCCGGCTG	407

RESULT	8	
AF195772		
LOCUS	438 bp	DNA linear BCT 01-NOV-2000
DEFINITION	Streptomyces netropsis SsgA (ssgA) gene, complete cds.	
ACCESSION	AF195772	
VERSION	AF195772.1	GI:11066162
KEYWORDS		
SOURCE	Streptomyces netropsis.	
ORGANISM	Streptomyces netropsis	
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE	1 (bases 1 to 438)	
AUTHORS	van Wezel,G.P., Rousseau,C. and Kraal,B.	
TITLE	Cloning and sequencing of the Streptomyces netropsis ssgA gene	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 438)	
AUTHORS	van Wezel,G.P.	
TITLE	Direct Submission	
JOURNAL	submitted (18-OCT-1999), Biochemistry University of Leiden	DO NOT

```

FEATURES
  source
    1. .438
      /organism="Streptomyces netropsis"
      /db_xref="taxon:55404"
    1. .438
      /gene="ssgA"
    1. .438
      /gene="ssgA"
      /note="similar to Streptomyces griseus SsgA; possible cel
division protein"
      /codon_start=1
      /transl_table=11
      /product="SsgA"
      /protein_id="AAG38483.1"
      /db_xref="GI:11056163"
      /translation="MRESVQAEVLMSFLVSEELSPFKIPVELRYETRDPAVVRMTFHL
GDAPVTWAFQRELLLDGINRPSGDDVHIATPTDGLSDVSIRLQVGADRALFRAGA
PLVAFTIETPKSVDLCEQFTIGREDSIAALCKLTAEQNC"

```

```

ORIGIN
Query: Match
      61.1%; Score 248.6; DB 1; Length 438;
Best Local Similarity 75.7%; Pred. No. 2.4e-22;
Matches 308; Conservative 0; Mismatches 99; Indels 0; Gaps 0

QY 1 atgagcttcctcgctcggagaagactcctctccgtattccggtggagctcggtacgag 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 ATGAGCTTCCTCGTCTCCGAGGAGCTCTCTTCAAGATCCCACTGCGAATCGGATACGAG 90

QY 61 acctgtgatacctacgcgcgtcggtgaccttctcatctgcgcggagatgccccggtgacc 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 ACCCGGGATCCCTACGCGGTGCGGATGACCTTCCACCTCCCGGGAGACCGCGCTGTGACC 150

QY 121 tggcgcttcggcgcgaggattgtctcatcacagagatccgcgcgacctcgcgggacggggac 180

```



[illegible]

RESULT	9
AX007220	
LOCUS	AX007220
DEFINITION	Sequence 5 from Patent WO0000613.
ACCESSION	AX007220
VERSION	AX007220.1
KEYWORDS	GT:9995086
SOURCE	.
ORGANISM	Streptomyces albus G. Streptomyces albus G Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 407)
REFERENCE	Kraal,B., Luiten,R.G. and Van Wezel,G.P. Reducing branching and enhancing fragmentation in culturing filamentous microorganisms
AUTHORS	PATENT: WO 0000613-A 5 06-JAN-2000;
TITLE	UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA (NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS VAN (NTL)
JOURNAL	

```

FEATURES             Location/Qualifiers
     source            1..407
                        /organism="Streptomyces albus G"
                        /db_xref="taxon:1962"
     gene              1..405
                        /gene="ssgA"
     CDS               1..>405
                        /gene="ssgA"
                        /note="Strain ATCC of Streptomyces albus G"
                        /codon_start=1
                        /transl_table=11
                        /product="SsgA"
                        /protein_id="CAC07386.1"
                        /db_xref="GI:9995087"
                        /translation="MSFLVSEELAFRIPVELRYETVDPYAVRLTHPLPGDAPVTWVFG
                        RELVGVLDAAAGDGVRCVPVGQTATREVEHITLQVSGEQALFRVGKAPLLAFLDRTD
                        QSLSGSERAHADFDOSHLDLALNRSLAEQSAG"
57 a 150 c 139 g 61 t

```

Query Match	59.9%	Score 243.8;	DB 6;	Length 407;
Best Local Similarity	74.9%;	pred. No. 9.7e-22;		
Matches 305;	Conservative	0;	Mismatches 102;	Indels 0;
				Gaps 0;

Qy	1	atagattctctcgttcgcgaagaactctcctcgtattccggtgagctgcgttaacgag	60
Db	1	ATGAGCTTCCTCGTCCGAGGAGCTCGCTCCGATCCGCTGGAGCTGCGGTACGAG	60
Qy	61	acctgtgatacctacgcctgcgcgtgcgaccttctatctgcccgagatgccccggtgacc	120
Db	61	ACCGTGTGATCCGTACCGGCTGCGGCTGACGTTCACCTCCCGCGAGACGCCCGGTACC	120
Qy	121	tggcgcttcggcgggagtgctgtcattcgcagcgaggtctcgcggcgctgcggggagcggggac	180

[illegible]

RESULT	10
AF195771	
LOCUS	566 bp DNA linear BCT 01-NOV-2000
DEFINITION	<i>Streptomyces albus</i> SsgA ( <i>ssgA</i> ) gene, complete cds.
ACCESSION	AF195771
VERSION	AF195771.1 GI:11066160
KEYWORDS	.
SOURCE	<i>Streptomyces albus</i> .
ORGANISM	<i>Streptomyces albus</i> Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1. (bases 1 to 566) van Wezel,G.P., Rousseau,C. and Kraal,B. Cloning and sequencing of the <i>Streptomyces albus ssgA</i> gene Unpublished JOURNAL REFERENCE 2 (bases 1 to 566) AUTHORS van Wezel,G.P. TITLE Direct Submission Submitted 18-OCT-1999 Biochemistry University Of Leiden. PO Box

**FEATURES**  
**source**  
1. .566  
/organism="Streptomyces albus"  
/db\_xref="taxon:1888"  
16. .433  
/gene="ssgA"  
16. .20  
/gene="ssgA"  
26. .433  
/gene="ssgA"  
/notes="similar to Streptomyces griseus SsgA; possible cell  
division protein"  
/codon\_start=1  
/transl\_table=11  
/product="SsgA"  
/protein\_id="AAG28482.1"  
/db\_xref="GI:11066161"  
/translat="MSFLVSEELAFRIPVELRYETVDPYAVRITFHLPGDAPVTWVFG

BASE COUNT	ORIGIN
75 a	219 c 192 g 80 t

	Query Match	59.9%	Score 243.8	DB 1	Length 566
	Best Local Similarity	74.9%	Pred. No. 8.9e-22		
	Matches 305	Conservative	0	Mismatches 102	Indels 0
Qy	1	atgagctcttcgtcgcggaagaactctcttcgtattccgtcggtgagctgcgttacgag	60		
Db	26	ATGAGCTTCCTCGTCTCCGAGAGCTGCCCTTCGCATCCGGTGGAGTCGCGTAGAG	85		
Qy	61	acctgtgatccctacgcgctgcggtcgtacactttcatctgcgccgagatgcgcccggtgacc	120		



positional base preference and amino acid composition.  
 Contains probable N-terminal signal sequence and  
 appropriately positioned PS00013 Prokaryotic membrane  
 lipoprotein lipid attachment site, so may be a  
 lipoprotein"  
 /codon\_start=1  
 /transl\_table=11  
 /label\_SCE19A.03c  
 /product="putative secreted protein"  
 /protein\_id="CAB50984.1"  
 /db\_xref="GI:5531352"  
 /translation="MAASRHLRRTTAVATVATIALTAGLTGCDVADKALDCVRT  
 ADATADSVTELOQAVENADPTQWEESLNSIDKDLRIGDQTNTDVKAVADDLGKAV  
 DNVRTSVENGDETDLSPFVTDAAAGELTKVCTP"  
 /gene="SCE19A.03c"  
 complement(1201..1233)  
 /note="PS00013 Prokaryotic membrane lipoprotein lipid  
 attachment site"  
 complement(1454..2191)  
 /gene="SCE19A.04c"  
 /note="rph"  
 complement(1454..2191)  
 /gene="SCE19A.04c"  
 /note="SCE19A.04c, rph, probable ribonuclease PH, len: 245  
 aa; highly similar to many e.g. SW:RNPH\_ECOLI  
 (EMBL:X00781), rph, Escherichia coli ribonuclease PH (238  
 aa), fasta scores; opt: 916 z-score: 1116.8 E(): 0, 60.8%  
 identity in 240 aa overlap. Contains Pfam match to entry  
 PF01138 RNase\_PH, 3' exoribonuclease family and PS01277  
 Ribonuclease PH signature"  
 /codon\_start=1  
 /transl\_table=11  
 /label=rph  
 /product="putative ribonuclease PH"  
 /protein\_id="CAB50985.1"  
 /db\_xref="GI:5531353"  
 /translation="MSRIDGGRTPOOLRVPIERGWSKHAGSVLVSPGDTKVLGNASV  
 TEGVPWRKSGVLEGTGVAEYGLPRATYZNRGDSRVKGRIGGTHSRIGSLRSLAV  
 IDYKALGENTVGLDCVLDQADGGTFTAAITGAYVALADAVAWAQGRKLKANRKPITG  
 TVSAVSGIVDGTPLLDLRYEDVDATDMNVVCTGDRGVQGTAAEPFAARDELN  
 TLLDLAVAGCTELAEQLKALDALTLE"  
 complement(1511..2185)  
 /gene="SCE19A.04c"  
 /note="Pfam match to entry PF01138 RNase\_PH, 3'  
 exoribonuclease family, score 266.70, E-value 3.1e-76"  
 complement(1805..1843)  
 /gene="SCE19A.04c"  
 /note="PS01277 Ribonuclease PH signature"  
 complement(2268..2501)  
 /gene="SCE19A.05c"  
 complement(2268..2501)  
 /gene="SCE19A.05c"  
 /note="SCE19A.05c, hypothetical protein, len: 77 aa;  
 unknown function, similar to parts of many sugar permeases  
 e.g. TR:P96159 (EMBL:U65013), malX, Vibrio furnissii PTS  
 (phosphopolypurvate-dependent sugar phosphotransferase  
 system) permease for glucose (523 aa), fasta scores; opt:  
 217 z-score: 287.9 E(): 1.1e-08, 48.1% identity in 77 aa  
 overlap. Contains Pfam match to entry PF00367 PTS\_EIIB,  
 phosphotransferase system, EIIB"  
 /codon\_start=1  
 /transl\_table=11  
 /label\_SCE19A.05c  
 /product="hypothetical protein"  
 /protein\_id="CAB50986.1"  
 /db\_xref="GI:5531354"  
 /translation="MASKAKEIVAGLGIDNIDEIGCTIRLRTEVNDPALVNEALK"  
 AAGHGVVKMGTAQVIGTGTADPIAAEIDMM"  
 complement(2385..2489)  
 /gene="SCE19A.05c"  
 /note="Pfam match to entry PF00367 PTS\_EIIB,  
 phosphotransferase system, EIIB, score 43.00, E-value  
 8.1e-11"

```

gene      2660. .3955
          /gene="SCE19A.06"
CDS
          2660. .3955
          /gene="SCE19A.06"
          /note="SCE19A.06, possible PTS transmembrane component,
          len: 431 aa; similar to many PTS
          (phosphoenolpyruvate-dependent sugar phosphotransferase

Query Match      27.1%; Score 110.4; DB 1; Length 35284;
Best Local Similarity 55.9%; Pred. No. 1e-05;
Matches 210; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 2   tgaagttcctgtctcggagaagaactctccttccgtattccggttgagctgcggttacgaga 61
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20142 TCAGGCTCGTCCGTGCCCGAGAGAGCGGCAFTCCGGTGC CGGCCGCTCGGCTACCACA 20201

QY 62   cctgtatccctacgcgcgtgcgcctgaccttctatctgcgcggagatgcgcccggtgacct 121
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20202 CGGACGATCCCTAGCCGCTGCACATCACCTTCCACATCGACTCCGGCCACCCGGTGCACT 20261

QY 122   gggcgtcggcgggagtggtctcatctacgagaggtcgcgcgccctgcgggacggggacg 181
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20262 GGACGTTCCGCCGGGACCTTCTGTGTGAGGGCGCTTCCGGCGCTCCGGGCACGGGACG 20321

QY 182   tccacatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 241
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20322 TGC GGCGTGTGGCCCTCGAAGACGAGGGCCGACGCTGTACTCGTCCGCTGAGCAGCC 20381

QY 242   ggaacgaccaggagatgttcgggctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 301
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20382 CGGACGGCGAGCCCTCCTGTGAGGCGCCACGCCGCCAGGTTCGGCTTGGCTGAGCGGA 20441

QY 302   cggacaagatctgtccgcctggcgaggagcgttccctctgcgcacttcgacccctgctcg 361
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20442 CCCTGCGGGCGGTGCCCGGGGACCGAGGCGCGAGCTCGGGATCGACGACGGGCTGG 20501

QY 362   acgaggcgtggaccg 377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20502 CCGAGCTGCTCGCCAG 20517

RESULT 12
SCL2
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS

SOURCE
ORGANISM
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS
REFERENCE
AUTHORS
TITLE
JOURNAL

```

David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid L2.

# FEATURES

```

source
1..38640
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid L2"
1..989
/gene="SCL2.01"
1..989
/partial
/gene="SCL2.01"
/note="SCL2.01, hypothetical protein, len: >328 aa; similar to TR:033236 (EMBL:298209) Mycobacterium tuberculosis hypothetical 49.8 kD protein MTC1174.11, len: 450 aa; fasta scores: opt: 843 z-score: 874.7 E(): 0; 45.9% identity in 296 aa overlap. Contains possible coiled-coils region"
/transl_table=11
/product="hypothetical protein SCL2.01"
/protein_id="CAB70913.1"
/db_xref="GI:6822207"
/transl_table=11
/translation="SWQAGSPREALYFERKYEGVLVWIGLLEKRVKTTDLAKDAQT
AVDHLREQVDHVAHDLALRDOLVALVETREERKQRAKQDEAGKALV
AEAEELASDQWRNAGERLSVDTWGLPLDRKSDDELHWRFSHARSASKRKQH
FALDAQREARIRKIRLVSEALNSDWDGTPAARYRDLMSWKWAKRAQRAQREHDD
LNNRFQAQDVFFAARSVFAERDAEQSENKLKEELYEAELKPLVPTDLKSARAAFR
SYNERWEAIGHPRDARPKSRGGCTRSSGSPRRRPPRSAGGTPRHRVPRA"
1010..1150
/gene="SCL2.02"
1010..1150
/note="SCL2.02"
/note="SCL2.02, unknown, len: 46 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCL2.02"
/protein_id="CAB70914.1"
/db_xref="GI:6822208"
/translation="MDKURSQVEQRAQGNDAKADKLARELEGRQALLDQALKGLHEF
GG"
complement(1218..3934)
/note="previously sequenced region SW:SCAPTRELA
EMBL:X87267 S.coelicolor apt & rela genes"
complement(1229..3772)
/gene="relA"
complement(1229..3772)
/gene="relA"
/note="SCL2.03c, rela, GTP pyrophosphokinase, len: 847 aa; identical to previously sequenced SW:RELA_STRCO (EMBL:X87267) Streptomyces coelicolor GTP pyrophosphokinase (EC 2.7.6.5) RelA, 847 aa and highly similar to SW:RELA_STRAT (EMBL:AF072829) Streptomyces antibioticus RelA protein, 841 aa; fasta scores: opt: 4969 z-score: 4906.2 E(): 0; 90.3% identity in 848 aa overlap and to SW:RELA_CORGL (EMBL:AF038651) Corynebacterium glutamicum GTP pyrophosphokinase (EC 2.7.6.5) RelA, 760 aa; fasta scores: opt: 1613 z-score: 1594.3 E(): 0; 58.7% identity in 774 aa overlap. Contains Pfam match to entry PF01842 ACT, ACT domain and match to Prosite entry PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="GTP pyrophosphokinase"
/protein_id="CAB70915.1"
/db_xref="GI:6822209"
/translation="MPDEAOLPLTAAPKESASAAKAPAPSAPOAKNDTHGPIQHAPAA
PVDPAEQOPRPKPLPAPRONAPVRAPAGOPARSGSSNRVRARLARLIGVQRANPYN
VLEPLRLIVRGNDPKIETSTLQJERAYQVAERHGRGKRKSGDPYITHPLAVTILL
AELGMDPATLMAGLHDJVEDTEYGLDRLDFGVDTLLVDGVTKDKRKGGEAAQA
ETVRKVMAMAKDPRLVTKLADRLHNMTRYLREKQEKRAKRETIYAPLAHRLG
MTIKWELEDLAFALPKMYDEIVLVAERAPKRDEYLAIVTDEVQDLRAARIKAT
VTRPKHYSVYQKIMVIRGRDFAEIVDLVIRLVDTVRDCYAAALGTVHARNKVPGR
FKDYTAMPKNMYQSLHTTVIGPGKGPVLEQLRTFDMHRAEYVAAHWKYQOEAVAG
ASKYRTDAPKSSGSKDHLNDMAHLRQLLDQKETEDEPGEFLESRLDLSNEVVF
TPKGDVIALPAGATPVDFAYAVHTVEGHTIRGARVGRNRLVPLESLINDGDLVEVFTSK
AAGAGPSRDWLGFKVSPRANKIRAMFESKERDEALEQGDALVRMKNQNPIDRIIL
TGDSLTLAHENRYSDISALYAAIGEHVSAPNIQVKLQALGGEAAFEEDSESVPP
SRGRKRRANADPGVWVKGVDVWVKLARCCTPVGDPPIIGFVTRGSGSVHRSDCV
NVDLSRPERITLEVEMPTOSSVFLVAIQVEALDRSLSDVTRVLSQDHVNLISA
VQTSRDRVATSRFTEMGDPKHLGHVKNVRGVEGYDYVRYTSARRPS"
complement(1256..1474)
/gene="relA"
/note="Pfam match to entry PF01842 ACT, ACT domain, score 63.10, E-value 6e-15"
complement(2378..2401)
/gene="relA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(3273..7460)
/note="previously sequenced region SW:SCSECAPT EMBL:X85969 S.coelicolor secD, secE & apt genes"
complement(3778..3781)
complement(3904..4568)
/note="previously sequenced region SW:SCAPTRELA
EMBL:X87267 S.coelicolor apt & rela genes"
complement(3956..4453)
/gene="apt"
complement(3956..4453)
/gene="apt"
/note="SCL2.04c, apt, adenine phosphoribosyltransferase, len: 182 aa; identical to previously sequenced SW:APT_STRCO (EMBL:X87267) Streptomyces coelicolor adenine phosphoribosyltransferase (EC 2.4.2.7) Apt, 182 aa and highly similar to SW:APT_ECOLI (EMBL:M14040) Escherichia coli adenine phosphoribosyltransferase (EC 2.4.2.7) Apt, 183 aa; fasta scores: opt: 600 z-score: 682.8 E(): 1.2e-30; 54.8% identity in 168 aa overlap. Contains Pfam match to entry PF00156 PriBosyltran, Phosphoribosyl transferase domain and match to Prosite entry PS00103 Purine/pyrimidine phosphoribosyl transferases signature"
/codon_start=1
/transl_table=11
/product="adenine phosphoribosyltransferase"

```

## COMMENT



JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351  
2 (bases 1 to 32704)  
Saunders,D.C. and Harris,D.  
Unpublished  
3 (bases 1 to 32704)  
Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.  
Direct Submission  
Submitted (03-AUG-2000) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

## COMMENT

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid 8A11.

## FEATURES

Location/Qualifiers  
1..32704  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 8A11"  
misc\_feature  
1..100  
/gene="SC8A11.01"  
/note="nominal overlap with Streptomyces coelicolor cosmid St9A4"  
1..2254  
/gene="SC8A11.01"  
1..2254  
/gene="SC8A11.01"  
/note="SC8A11.01, possible transcriptional regulator (fragment), len: >750 aa; similar to TR:Q9S124 (EMBL:AI109747) Streptomyces coelicolor probable transcriptional regulator SCJ21.13, 919 aa; fasta scores: opt: 818 z-score: 913.7 E(); 0: 32.6% identity in 775 aa overlap. Contains Pfam match to entry PF00196 GerE, Bacterial regulatory proteins, luxR family and match to Prosite entry PS00622 Bacterial regulatory proteins, luxR family signature. Also contains a possible helix-turn-helix motif at residues 699..720 (+3.25 SD)"  
/codon\_start=2  
/transl\_table=11  
/product="putative transcriptional regulator (fragment)"  
/protein\_id="CAC01573.1"

gene  
CDS

/db\_xref="GI:9716212"  
/translation="IGAARTSSPGTYTLGTHAEVLPLESPSAAALDADAAGIADA  
VRILQRAAGNPLAVELPRAAGISPLDLDPLTQLETAFASTRDLSLRECRFTFL  
LVLAEPAPLNQLLDVASRLAGSEVTVYAEQAVDAGLVLTGTPTPEHPMBSAI  
YTRATVPDLRLSTHRLAETLGSGPGRRLVHLAAATLGPDDDELQGLERFADDAQRGQ  
LAAAPVALRQAGELVHPRRQTGLLVRAAEALASEINDRQAQILLOGLLRPAEPGTER  
ARMLVSKDAAFEPDQPRQIDMAAGAFDVGSTVAENLWRAARAARCFDQGD  
RVRAQAAALDRWKPDPAHVLTVRAYTEPYRRGTDLIALEKLRDREGRLLHL  
GSGMAIGDVGTRATYLAQAASVWRSGRLGLLARSLAGSNPRLYLGLQAQARESAE  
GIALAEFGEMIVMLGKATSLTAVLRGEAARSVRELRAHLSLFPVMPASVMAQ  
QVEGLLAFDSRAVEYDALARFDTDPHSHSRWLLYLPDLVDAAAAGNEQARE  
LLVELPELARLPDSEMMIVARTYSTAVLAPDQDAEDCYDSALSALPDTWPLARLHL  
QHGRLLRRQRNDKPLRLARDEFDRVGAQPMADAREQLRAAGESDGRRRPSKGE  
SLTVQEROIAELASQGLSNREITQGRFLFSHRTVGAHLYRIYPRLGITSRGRLSAALA  
LGDDQPTSGG"  
2042..2239  
/gene="SC8A11.01"  
/note="Pfam match to entry PF00196 GerE, Bacterial  
regulatory proteins, luxR family, score 49.10, E-value  
9.7e-11"  
2093..2176  
/gene="SC8A11.01"  
/note="PS00622 Bacterial regulatory proteins, luxR family  
signature"  
complement(2184..2453)  
/gene="SC8A11.02c"  
complement(2184..2453)  
/gene="SC8A11.02c"  
/note="SC8A11.02c, unknown, len: 89 aa"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC8A11.02c"  
/protein\_id="CAC01574.1"  
/db\_xref="GI:9716213"  
/translation="MASAATAGTAGALPVMGQAHADTGRGLPSLMDVDRSVVNPENA  
ITVTVDQVRAEVAQLMGATEEETPRSGRLRAPRACRGW"  
2787..3257  
/gene="SC8A11.03"  
2787..3257  
/gene="SC8A11.03"  
/note="SC8A11.03, possible regulator, len: 156 aa; similar  
to TR:Q9S2F7 (EMBL:AL096852) Streptomyces coelicolor  
putative regulator SCE19A.24, 142 aa; fasta scores: opt:  
272 z-score: 353.0 E(); 3.4e-12; 36.9% identity in 130 aa  
overlap"  
/codon\_start=1  
/transl\_table=11  
/product="putative regulator"  
/protein\_id="CAC01575.1"  
/db\_xref="GI:9716214"  
/translation="MSGDHGVQAOHSAQAALLPLSLCSQMTGALEWEDVPAEPRYD  
PDHLLVTRFAPGAPVTWVGFDLLHGLRTTSGIGDYQVWADTPTDRETALQV  
NAHGDIATFSLPVPPELEWIDRTYLVHPAGTSSRLGTDAFLSKLDFDEPSSR"  
complement(3297..7058)  
/gene="SC8A11.04c"  
complement(3297..7058)  
/gene="SC8A11.04c"  
/note="SC8A11.04c, probable secreted peptidase, len: 1253  
aa; similar to TR:Q9RL54 (EMBL:AL121596) Streptomyces  
coelicolor probable secreted peptidase SCF51A.10, 1245 aa;  
fasta scores: opt: 1827 z-score: 1848.3 E(); 0: 37.2%  
identity in 1300 aa overlap, to TR:P95684 (EMBL:D83672)  
Streptomyces albobogiseolus subtilisin-like protease, 1102  
aa; fasta scores: opt: 1365 z-score: 1380.6 E(); 0: 33.4%  
identity in 1268 aa overlap and to Streptomyces coelicolor  
probable secreted peptidase SC8A11.16c, 1239 aa; fasta  
scores: opt: 1835 z-score: 1467.4 E(); 0: 46.9% identity  
in 1279 aa overlap. Contains Pfam matches to entries  
PF00082 Peptidase\_S8, Subtilase family and PF02225 PA, PA  
proteases, subtilase family, aspartic acid active site,  
PS00137 Serine proteases, subtilase family, histidine  
active site and PS00138 Serine proteases, subtilase

family, serine active site. Also contains possible N-terminal region signal peptide sequence"

/codon\_start=1  
/transl\_table=11  
/product="putative secreted peptidase"  
/protein\_id="CAC01576.1"  
/db\_xref="GI:9716215"  
/translation="MPIPWRSGLTVGTAALVLTATPAASALPGTVPAPPTAPA  
GPRVLTLLGDRVTTGDEGAFVLSVTDPHRSGGAHMTVGSYVYVPAAPVYL  
GSGALDERLNFNTELEDGYDDARADELPVITYTDAARS LGARPEGARFRALSS  
IRGAALSABESRAADFTWLSLTGDAAGSAAARSATSGRLAGGIAKVLWDGKRAT  
LSDTTAIGADPVMSGGTGGVGVAVLDITGVNDAGHPDFAGRIATASFVDDVDTR  
NKGHTVATAGTGAAGVEKVGASLHIGKVLDSNGSQDQSVLWAGMEWAYRD  
QKAKIVMSLGSPTDGTPLSRVAVNLAEATGFAVAGNSGPAYTVGTAAADA  
ALTGVAVGPKGVLDLADFSRRPVRGDNVAVKPDLTAPGVGLAARSRYAPGEGAY  
QSLSGTSLASGPAVAGAAALLAAPHDPWTGRLKEALVGTACTQRESPDAGSRVDV  
AAVRSTTSLAGDAFAAHYPTPGQTVRVDVYTNNGFAPVADLIALSPAELPEGLF  
TUSEAQTVPAHGTA SVGTHLDAEDNGAVATRLVAGDAGVIALPSPVKEGR  
RATLALTAKDHDHPLSGTVILKDVERTAPKYSVDASGRDLRLSPSTYSVMNSA  
VPGVDCGTHLGFAMFTAPEVVLADATVAFDADDLKAAATPRATANGFLRIDQVRG  
NTGLFPFMSYVAEYWRVDSLWVTPPEVTRGTSYTPATRWQIQPPLTFSAGSQTFDD  
VTQSRSPQLPECTRAYRAVWAGDGSATERGAIEVRVAVVRSSTVAPTDQAAAAE  
KAGARQLLINDGIGFDEWADLPEAPLIPFVAGSIEDSDSARLLAREGAGTTTLRVVS  
HPVRYAYDLVRHGDVAPRDPSPYRPAPELARVDVDTFRDTSQRAVEYRQDLSLIGQ  
PLGIVTVQRAQCELTSWTADDDVRWVSFASRDLQCGRVARSYPRSTRTTFEAP  
IOHPRLLSONGTSGCGFPFRAGDNISTVMTAWGDSGHAGVWADGTSRISLYOGGE  
LIGEDYNERIVVMVGLSPGPKPVRLVLEGRNLPDRPYSTRFTVNDFTSATDPTRL  
TPLPLVQLDYAVAVDLSGRHRTTELTVTASHLEGAAGAGAIRTAIVEVSYDDGATWH  
RTALRSAGDGTARLADPGARVARSUAKDTFEGNGVGTGLIRAFGLR"  
misc\_feature  
complement(4374..4613)  
/gene="SC8A11.04c"  
/note="Pfam match to entry PF02225 PA, PA domain, score  
19.00, E-value 0.058"

Query Match 19.28; Score 78; DB 1; Length 32704;  
Best Local Similarity 51.4%; Pred. No. 0.11;  
Matches 180; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 38 ttccggtgagctgcgttacgagactgtatccctacgcgctgagctgacatttcac 97  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 2896 TTCCGGGGAGTTCGGTTCATCCGACACACCCCTCTCGTCACGATCGCTTCGCC 2955  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 98 tgcccgagatgccccggtgacctggcggttgcggcgaggagttgctcatcgacggaggtc 157  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 2956 CCAGAGGCGCCCCACCGTACCTGGATGTGCGCGCTGACCTGCTCACGAGGCGCTGC 3015  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 158 cgcgccgctgcgggagcgggagtcacatcgcccgccgacccgagagcgttcggcg 217  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 3016 GCACACGAGCGGCTGGGGAGCTCCAGGTGTGGGCGCACACCCCGACCGCGGAGA 3075  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 218 aggtcctgacccgctgcagtgaggagagcagcagcagtgatcttcgggtcgccagcgccg 277  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 3076 CCGGTGCTCCAGGTCAACGCACACCGGTGACATAGGCATCTTCAGCGTCCGCCCTCCCG 3135  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 278 cgctggtgctctctctgacgcagcagcagcagcagcagcagcagcagcagcagcagc 337  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 3136 AACTGGAGGAGTGATGACCGCACGTACCTGACGTCCCGCGCGGAGACGAGTCTGTCG 3195  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 338 tcgcccagcttcgacgctgctgacgagcagcagcagcagcagcagcagcagcagcagc 387  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 3196 GGCTCGGAACCGAGCGCTTCTCTCCAAGCTGTTCGACGAACCGAGGCGC 3245  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15  
SCH22A/c  
LOCUS SCH22A 29625 bp DNA linear BCT 07-MAR-2000  
DEFINITION Streptomyces coelicolor cosmid H22A.  
ACCESSION AL159178  
VERSION AL159178.1 GI:7210989  
KEYWORDS ABC transporter ATP-binding protein; cold-shock protein; DEAD-box RNA helicase; epimerase; f40; glycosyltransferase; histidine kinase; integral membrane protein; lipoprotein; membrane protein; oxidoreductase; regulatory protein; response regulator; secreted

SOURCE ORGANISM  
REFERENCE AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics.  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid H22A Lies between and overlaps with cosmids H35 and H63 on the AseI-H genomic restriction fragment.  
Location/Qualifiers  
1. .29625  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid H22A"  
1. .1399  
/gene="SCH22A.01"  
1. .1399  
/partial  
/gene="SCH22A.01"  
/note="SCH22A.01, possible regulatory protein (fragment), len: >465 aa; similar to TR:CAB59599 (EMBL:AL132662) Streptomyces coelicolor hypothetical regulatory protein SCrll1.22, 877 aa; fasta scores: opt: 708 z-score: 765.1 E(): 0; 34.5% identity in 466 aa overlap. Contains Pfam match to entry PF01590 GAF, GAF domain and possible

## FEATURES

## source

## gene

## CDS



```
coiled-coil region at residues 48..60"
/codon_start=2
/transl_table=11
/product="putative regulatory protein (fragment)"
/protein_id="CAB76970.1"
/db_xref="GI:7210990"
/translation="EGPSGPAQSGGPAAPAGRGDDPEYASLLDPLSSDLNRIG
EQHALARQRTQLLEQAVNITGETLELPVAVLRVIRVTAMDVGARYGAMVLQDEG
RLLEFIPGLTKELADLEGVELPRGLGLHLHPEPLRVKDISRHPEPAGPPG
HPMRSLLGVAISVGRYGNLYLSERKDGQPFDRHDEGVIRALAGTAGVATENARLY
QVRRNSGQKLLPLRPDLNPLFTAGAIRPASAPAGVGWDVMDMLPGGACAAVI
GDVGHDRRAADMDSQIRNMLRALLYDPGAPSTSLARDTMMNALDEAPVATALLA
RLEPADGTWQLRWSAGHPPLVLLPDGVRVLDMEPGLGVADLPDRPDHCRPLPV
DSTVILFDGLVEVFGQPLRDGLDALAVTASRLVGLAPEDLCHALVAHRPNGHDDKA
VIALRTPLTYSVGS"
misc_feature
1..130
/feature="SCH22A.01"
/feature="nominal overlap with S. coelicolor cosmid sth35"
215..652
/feature="SCH22A.01"
/feature="pfam match to entry PF01590 GAF, GAF domain, score
89.60, E-value 6.5e-23"
complement(1435..1917)
/feature="SCH22A.02c"
complement(1435..1917)
/feature="SCH22A.02c"
/feature="SCH22A.02c"
/feature="SCH22A.02c, hypothetical protein, len: 160 aa;
similar to TR:Q9X7Q5 (EMBL:AL049587) Streptomyces
coelicolor hypothetical 17.1 kD protein SC5F2A.32, 161 aa;
fasta scores: opt: 256 z-score: 310.3 E(): 7.3e-10; 39.4%
identity in 160 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCH22A.02c"
/protein_id="CAB76971.1"
/db_xref="GI:7210991"
/translation="MLGPRVGRQPRTPFGCDGMSTLLQPMPHRVLTLPAPPAVRL
ARETAQALAEWGNPGRPAVAPALLILSELVNSVRHASPSEQVTVTYAAGDDCLA
FAVDRHPHQPRLHCARTGRTGGLATVWELTGLGTAIVRGGDGRGKSIWTLPL
"
complement(1969..3258)
/feature="SCH22A.03c"
complement(1969..3258)
/feature="SCH22A.03c"
/feature="SCH22A.03c"
/feature="SCH22A.03c, possible membrane protein, len: 429 aa.
Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB76972.1"
/db_xref="GI:7210992"
/translation="NAGATLAFVLVAETAAARYGLPGMTNQAKELVPAPASGPLLY
AGLALTMVLTWRQLRVAANAVGVDLTVALVRWAADATASSHSEFGNALWVLGCG
VIALRTGPERILLKGLVGLLVAGRTKGDWLLITSKTRPDVLPYVATADHAL
GNPSLWLGATPAGVGEHILDVYQVLAFAVVAALYQLRVAAERFRPHRLVRT
FLILGLGPAIYVFPVGPVAYGTGTGAEWAIALWLPHTLPVPPHPTFYD
VTPRNCPSLHTNATVIFHSRKGPRVLRWAGVFLVATLTLGFCYHVAIIDLIA
GVVEATVPEAGRLSDRGWRSGLLVAGALVFTAILASTRYLSLEARNPWFVGPL
LLAMASVTHGVYRTKSWEPVTAAPALPEPRLEAA"
complement(3337..3957)
/feature="SCH22A.04c"
complement(3337..3957)
/feature="SCH22A.04c"
/feature="SCH22A.04c, unknown, len: 206 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCH22A.04c"
/protein_id="CAB76973.1"
/db_xref="GI:7210993"
/translation="MNSIDSVTLVADTEAARFVADAFGLDGSVRVLRASDDPTSGF
KGFTLSLVQAQGNVDALFTAAVDAGATVLPKPAKSLWGVGVVRAPDGTVMQIATSA
KADTAPVTRDNDVILLGVEDVKATKRFYVEQGLTVGKSGFKYVEATGFTGKLS
LYKRALAKVAGVSADGTGSHRLVVSGGTRPFADPDGFAWQPELSH"
RBS
gene
CDS
4072..4078
4088..4672
/feature="SCH22A.05"
4088..4672
/feature="SCH22A.05"
/feature="SCH22A.05"
/feature="SCH22A.05, possible stress response protein, len:
194 aa; similar to TR:AAF11772 (EMBL:AF002055) Deinococcus
radiodurans tellurium resistance protein Terz, 195 aa;
fasta scores: opt: 300 z-score: 364.6 E(): 6.8e-13; 31.5%
identity in 178 aa overlap and to SW:TERZ_SERMA
(EMBL:L38824) Serratia marcescens tellurium resistance
protein Terz, 193 aa; fasta scores: opt: 263 z-score:
321.2 E(): 1.8e-10; 28.2% identity in 195 aa overlap"
/codon_start=1
/transl_table=11
/product="putative stress response protein"
/protein_id="CAB76974.1"
/db_xref="GI:7210994"
/translation="MITLTKEGGPADLDGVTHLSIGVSWDPTAGSGGVLKLRRTG
TDLDLIAVAMOGDPVRLAGLDSDPMGNSLLHSGDNQTHGDDGDETVTVEFARLP
SAITSIVFAAAYKKGSSFOKARNISFKYVDATGGSSEQVADIMPILLSQDNGCAVAK
AVRVGGTWKLEVNETGKIKOGDEHALMRFAYSK"
complement(4798..5529)
/feature="SCH22A.06c"
complement(4798..5529)
/feature="SCH22A.06c"
/feature="SCH22A.06c, possible membrane protein, len: 243 aa.
Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB76975.1"
/db_xref="GI:7210995"
/translation="MGTDFATWTRFEDERRAAQCDPMWEGAVLHRVAVAGTOR
FOVGDDGDGANLVAAEAGADYARAVLFAVEQNRHARLLARLAAGDRPALSGHW
SDTVFRLRLGLLELTLVLMIAEVALRYRYRALRDGTDALTSEVAGRIILADERRH
VPFHGRHLHSLAELPAVTRPVVLMWRLLLLAATVVAADHGAGLRILGVGRRRFAA
DVMASASVAAEAVLTPRPAWSSGA"
complement(5539..5545)
/feature="SCH22A.07c"
complement(5537..6419)
/feature="SCH22A.07c"
complement(5537..6419)
/feature="SCH22A.07c"
/feature="SCH22A.07c, possible regulatory protein, len: 260
aa; similar to SW:BRPA_STRHY (EMBL:M64783) Streptomyces
hygroscopicus bialaphos biosynthetic pathway regulatory
Query Match 14.6%; Score 59.4; DB 1; Length 29625;
Best Local Similarity 54.9%; Pred. No. 23;
Matches 117; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 179 acgtccacatcgcccgccgacccgagagcgttcggcaggtccgtgacccgctcag 238
DB 28815 ACGTCTCACCCCGCCGCGGCGCTGCCCGCCGACGCCGCCGCGCGCGCGGC 28756
QY 239 tggggagcagcagcgatgttcggggtcgccgacgcgcgcgtggtggtcccttcctggacc 238
DB 28755 CCGGCGTCCGCGCGCGCTGAGCCTCTCTGACACGACGAGGTGCTGCTCGCGTCCGGGCG 28696
QY 299 gcacggacaagatcgtgcgcgtggggcagagcgttccttcgcgcgacttcgacgcctgc 358
DB 28695 GCGGCGAGACCTCGTTGACAGGGGACGCCGCCACCCAGGCGATCTCGGCGTCCGGCGCGAAC 28636
QY 359 tcgacgagcgctgacgcgcacatcctgagcgagg 391
DB 28635 TC CGGAAGGTGCAGGACCTGGAGCTGCGCGAGG 28603
Search completed: July 18, 2002, 11:22:25
Job time: 13755 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:36:31 ; Search time 646.57 Seconds  
(without alignments)  
1080.756 Million cell updates/sec

Title: US-09-749-185-6

Perfect score: 407

Sequence: 1 atgagcttctctgctcgga.....gaggagcagacgcggctg 407

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802.\*  
1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	407	100.0	407	21	AAZ49730 Streptomyces golde
2	259.8	63.8	407	21	AAZ49728 Streptomyces grise
3	259.8	63.8	438	21	AAZ49727 Streptomyces grise
4	248.6	61.1	407	21	AAZ49731 Streptomyces netro
5	243.8	59.9	407	21	AAZ49729 Streptomyces albus
6	58.8	14.4	45624	22	AAF88315 S. spinosa DNA fra
7	58.8	14.4	50000	22	AAF88312 S. spinosa DNA fra
8	54.4	13.4	1350	23	AA554328 Pseudomonas aerugi
9	54.4	13.4	2760	16	AAQ79921 P. aeruginosa accB

10	52.2	12.8	65140	22	AA17184 Streptomyces nous
11	52.2	12.8	125401	22	AA17186 Streptomyces nous
12	51.8	12.7	1879	12	AAO11126 Sequence encoding
13	51.6	12.7	109519	22	AA508693 Micromonospora DNA
14	51	12.5	53789	19	AAV21187 Amycolatopsis medi
15	49.4	12.1	38734	20	AAZ32020 Human METHI relate
16	49.4	12.1	38734	22	AAC90077 AL021529 cDNA clon
17	49.4	12.1	47981	22	AAF30757 Micromonospora meg
18	49.2	12.1	2743	22	AAH14184 Human cDNA sequenc
19	49.2	12.1	9210	22	AAH52046 Mycobacterium tube
20	49.2	12.1	4403765	22	AAI99683 Mycobacterium tube
21	48.8	12.0	1533	20	AAZ34147 Human PRO1072 nucl
22	48.8	12.0	1533	21	AAZ78533 Human PRO1072 (UNQ
23	48.8	12.0	1533	22	AA546015 Human DNA encoding
24	48.8	12.0	58857	21	AAA58471 Nucleotide sequenc
25	48.6	11.9	47981	22	AAF30757 Micromonospora meg
26	47.6	11.7	966	23	AA588065 DNA encoding novel
27	47.6	11.7	1218	23	AA587399 Sequence of macroc
28	47.6	11.7	1641	8	AAAT0466 Human protocadheri
29	47.6	11.7	3353	17	AAAT03572 S. clavuligerus cl
30	47.4	11.6	978	22	AA14506 Streptomyces clavu
31	47.4	11.6	11604	22	AA14501 Pseudomonas aerugi
32	47.4	11.6	13029	23	AA551470 S. clavuligerus cl
33	47.4	11.6	15079	16	AAQ91580 Streptomyces clavu
34	47.4	11.6	15079	22	AAQ14499 eryA region of S.
35	47.4	11.6	29879	14	AAO46806 Human herpesvirus
36	47.2	11.6	154746	24	AAO25519 Mycobacterium tube
37	47	11.5	4411529	22	AAI99682 Alcaligenes sp. Po
38	46.8	11.5	12019	20	AAI18867 Chimpanzee adenovi
39	46.8	11.5	36519	19	AAV22141 S. cellululosum DNA
40	46.8	11.5	37856	22	AAI11992 N. magadali bacter
41	46.8	11.5	48300	22	AA611281 Platenolide synth
42	46.6	11.4	44377	18	AA78508 Platenolide synth
43	46.6	11.4	44377	18	AA80414 Pseudomonas sp. WF
44	46.4	11.4	11279	21	AAA83889 Streptomyces nous
45	46.4	11.4	65140	22	AA17184 Streptomyces nous

ALIGNMENTS

RESULT 1  
AAZ49730  
ID AAZ49730 standard; DNA; 407 BP.  
XX  
AC AAZ49730;  
XX  
DT 18-APR-2000 (first entry)  
DE Streptomyces goldeniensis ssgA gene.  
XX  
DE ssgA: liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumor agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 21386;  
KW bioinsecticide; receptor agonist; antagonist; biomass; ds.  
XX  
OS Streptomyces goldeniensis.  
XX  
FH Key Location/Qualifiers  
FT mat\_peptide i..405  
FT /\*tag= a  
FT /label= SsgA\_protein  
XX  
XX WO200000613-A1.  
XX  
PD 06-JAN-2000.  
XX  
PF 25-JUN-1999; 99WO-NL00395.  
XX  
PR 26-JUN-1998; 98EP-0202148.  
XX  
PA (UYLE-) RIJKSUNIV LEIDEN.



Db	151	tgggcggttcggccgagctgtctgtcgacgggtctaacagaccccgagcggcgacggcgcat	210
Qy	181	gtccacatcgccccgcggccgacccgggagacggttcggcgaggtcctgatccgctcgagtg	240
Db	211	gtcacatcgccccgacccgagcggccctcgagagtgctccacatcgggtccaggtc	270
Qy	241	ggagcgacacgagcgatgttcgggttcggcgcacgagcggccgctagtgtgcttctgacccg	300
Db	271	ggcgcgacccgtgcgtgttcggggcgggagcgacccgctggtggtctctcgaccgg	330
Qy	301	acggacaaagatcgtccgctggggcgagcggttccctcgccgacttcgacgccctgctc	360
Db	331	acggacaaagcttctgcgcttcggcccgagcgacacgctgggtgacttcgacggcaacctg	390
Qy	361	gacgagggcgtgaccgcgcatcttcggcgaggagcagaacccggctg	407
Db	391	gaggacgactggcccgcatcttcgcccggagcagacgagcagacgacggcgctg	437
RESULT 4			
AAZ49731	ID	AAZ49731 standard; DNA; 407 BP.	
XX	AC	AAZ49731;	
XX	DT	18-APR-2000 (first entry)	
XX	XX	Streptomyces netropsis ssGA gene.	
DE	XX		
XX	XX	ssGA; liquid culture; filamentous bacteria; secondary metabolite;	
KW	XX	mycelium; antibiotic, antitumour agent; immunosuppressive agent;	
KW	XX	hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;	
KW	XX	herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23940;	
KW	XX	bioinsecticide; receptor agonist; antagonist; biomass; ds.	
OS	XX	Streptomyces netropsis.	
XX	Key	Location/Qualifiers	
FT	mat_peptide	1..405	
FT		/*tag= a	
FT		/label= SsgA_protein	
XX	XX	WO200000613-A1.	
PN	XX	06-JAN-2000.	
PD	XX		
XX	XX	25-JUN-1999; 99WO-NL00395.	
PF	XX		
XX	XX	26-JUN-1998; 98EP-0202148.	
PR	XX	(UYLE-) RIJKSUNIV LEIDEN.	
XX	XX	(NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.	
PA	XX	Van Wezel GP, Kraal B, Luiten RGM;	
XX	PI		
XX	XX	WPI; 2000-147269/13.	
DR	XX	P-PSDB; AAY44652.	
XX	XX		
PT	PT	Reducing branching and enhancing fragmentation in filamentous	
PT	PT	microorganisms used to improve their liquid culturing properties -	
XX	XX	Disclosure; Fig 5; 60pp; English.	
XX	XX		
CC	CC	The present sequence is a DNA (ATCC 23940) encoding S. netropsis SsgA	
CC	CC	protein. SsgA reduces branching and fragment septation and enhances	
CC	CC	fragmentation of mycelium in liquid culture resulting in lower viscosity	
CC	CC	of culture broths. Filamentous bacteria can be transformed with ssGA	
CC	CC	gene-containing plasmid to enhance the production of secondary	
CC	CC	metabolites such as, antibiotics, antitumour agents, immunosuppressive	
CC	CC	agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine	
CC	CC	agents, herbicides, antiparasitic agents, ruminant growth promoters,	
CC	CC	bioinsecticides, receptor agonists and antagonists and biomass.	
XX	XX		





Db 24831 gggagcttctgaagacacacctacgacggtgctgaacctggtgatgaacacgggtgatg 24890  
QY 343 gacttcgacgcctgctcgaacgagcgctggacgcatactg 384  
Db 24891 aacgtcgacgagtgctgagccagctgctcgcacaaactg 24932

RESULT 8  
AAS54328  
ID AAS54328 standard; DNA; 1350 BP.  
XX AC AAS54328;  
XX AC AAS54328;  
DT 13-FEB-2002 (first entry)  
XX Pseudomonas aeruginosa DNA for cellular proliferation protein #459.  
XX Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX Pseudomonas aeruginosa.  
XX WO200170955-A2.  
XX 27-SEP-2001.  
XX 21-MAR-2001; 2001WO-US09180.  
XX 21-MAR-2000; 2000US-191078P.  
XX 23-MAY-2000; 2000US-206848P.  
XX 26-MAY-2000; 2000US-207727P.  
XX 23-OCT-2000; 2000US-242578P.  
XX 27-NOV-2000; 2000US-253625P.  
XX 22-DEC-2000; 2000US-257931P.  
XX 16-FEB-2001; 2001US-269308P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI; 2001-611495/70.  
XX P-PSDB; AAU36469.  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX Claim 27; Seq ID No 7965; 51lpp; English.  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 1350 BP; 256 A; 468 C; 427 G; 199 T; 0 other;

Query Match 13.4%; Score 54.4; DB 23; Length 1350;  
Best Local Similarity 47.6%; Pred. No. 0.087;  
Matches 160; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 44' tggagctggttacgagacacgtgatccctacgcccgtggtgaccccttcctcgtcccg 103  
Db 824 tcgagttcctacgagacggccgcttctacttctacgagatgaacacctcgtgcagg 883  
QY 104' gagatgccccggtgacctggcggttcggggaggttgcctcatcgacggaggtccgcggc 163  
Db 884 tggagcaccgggtatctgagatggtcacccggtgctgacatcgtcaaggagatgctgcgca 943  
QY 164 cgtcggggagcgggacgtccacatcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 223  
Db 944 tcgctccggcgagagctctcgatccgcccagagagcgttgatcatccgcgcgcgcgc 1003  
QY 224 tgatcgctcgtcaggtgggagcgcacgagcgtgctccggggtcggcacggcgccgctgg 283  
Db 1004 tggagtcgggataacgcgcgaagacccgaagacattctatgccagcccgccgcaaggta 1063  
QY 284' tggccttcctgacccgacggacaagatcgtgcccgtgggagagcgttccctcgcgcg 343  
Db 1064 agcacttcacgcccccgcggaacggcggtgctgctgactcgacacctctacagcggt 1123  
QY 344 acttcgacgcccgtcgtcgcacgagggcgctggacgcga 379  
Db 1124 acagcgtgcccgaactacgactcgtggtcgga 1159

RESULT 9  
AAQ79921  
ID AAQ79921 standard; DNA; 2760 BP.  
XX AC AAQ79921;  
XX AC AAQ79921;  
DT 16-JUL-1995 (first entry)  
XX DE P. aeruginosa accB and accC genes.  
XX KW Acetyl-CoA-carboxylase; accB gene; accC gene; malonyl-CoA; lipid;  
KW oilseed; transgenic plant; biotin-carboxylase; BC;  
KW biotin carboxyl carrier protein; BCCP; ss.  
XX OS Pseudomonas aeruginosa PaO.  
XX FH Key Location/Qualifiers  
FT CDS 861..1334  
FT /tag= a  
FT /label= accB  
FT /note= "encodes the biotin carboxyl carrier  
FT preprotein"  
FT CDS 1349..2701  
FT /tag= b  
FT /label= accC  
FT /note= "encodes biotin-carboxylase"  
XX WO9429467-A.  
XX 22-DEC-1994.  
XX 06-JUN-1994; 94WO-US06447.  
XX 08-JUN-1993; 93US-0074121.  
XX (CALJ ) CALGENE INC.  
XX Best E, Knauf V;  
XX WPI; 1995-036492/05.  
XX P-PSDB; AAR66743; AAR66744.  
XX Increasing lipid levels in plants by modulating malonyl-CoA  
PT levels - partic. for oilseed crops, also nucleic acid









FT	RBS	complement (15369..15373)	FT	RBS	complement (40899..40902)
FT	FT	/*tag= v	FT	FT	/*tag= aw
FT	CDS	complement (15380..16414)	FT	CDS	complement (40887..41576)
FT	FT	/*tag= w	FT	FT	/*tag= ax
FT	FT	/product= "EvrB"	FT	FT	/product= "EvrU"
FT	CDS	complement 16419..17873	FT	CDS	complement (41679..42707)
FT	FT	/*tag= x	FT	FT	/*tag= ay
FT	FT	/product= "EvrC"	FT	FT	/product= "EvrV"
FT	CDS	complement (17870..18934)	FT	FT	complement (42714..42717)
FT	FT	/*tag= y	FT	FT	/*tag= az
FT	CDS	/product= "EvrD"	FT	FT	complement (42810..43799)
FT	FT	19374..20906	FT	FT	/*tag= ba
FT	FT	/*tag= z	FT	FT	/product= "EvrW"
FT	FT	/product= "EvrE"	FT	FT	complement (43807..43811)
FT	CDS	21064..22542	FT	FT	/*tag= bb
FT	FT	/*tag= aa	FT	CDS	complement (43799..44866)
FT	FT	/product= "EvrF"	FT	FT	/*tag= bc
FT	FT	21056..22542	FT	FT	/product= "EvrX"
FT	RBS	/*tag= ab	FT	FT	complement (45014..45760)
FT	CDS	22748..24172	FT	FT	/*tag= bd
FT	FT	/*tag= ac	FT	FT	/product= "EvrY"
FT	FT	/product= "EvrG"	FT	FT	complement (45767..45770)
FT	FT	22736..22740	FT	FT	/*tag= be
FT	RBS	/*tag= ad	FT	CDS	complement (45962..46714)
FT	CDS	complement (24177..25223)	FT	FT	/*tag= bf
FT	FT	/*tag= ae	FT	FT	/product= "Evr2"
FT	FT	/product= "EvrH"	FT	FT	complement (45952..45956)
FT	FT	complement (25230..25233)	FT	FT	/*tag= bg
FT	FT	/*tag= af	FT	CDS	complement (47156..49234)
FT	CDS	25550..26626	FT	FT	/*tag= bh
FT	FT	/*tag= ag	FT	FT	/product= "EvrA"
FT	CDS	26685..30479	FT	FT	51627..52715
FT	FT	/*tag= ah	FT	FT	/*tag= bi
FT	FT	/product= "EvrJ"	FT	FT	/product= "EvsB"
FT	RBS	26672..26676	FT	FT	51629..51622
FT	FT	/*tag= ai	FT	RBS	/*tag= bj
FT	CDS	/*tag= aj	FT	FT	52889..53557
FT	FT	/product= "EvrK"	FT	FT	/*tag= bk
FT	FT	complement (31885..31888)	FT	CDS	/product= "EvsC"
FT	FT	/*tag= ak	FT	FT	53554..54207
FT	CDS	complement (31941..32882)	FT	FT	/*tag= bl
FT	FT	/*tag= al	FT	FT	/product= "EvsA"
FT	FT	/product= "EvrL"	FT	FT	complement (54362..55117)
FT	CDS	complement (33167..34405)	FT	FT	/*tag= bm
FT	FT	/*tag= am	FT	FT	/product= "EvsB"
FT	FT	/product= "EvrM"	FT	FT	complement (55125..55128)
FT	RBS	complement (34414..34418)	FT	FT	/*tag= bn
FT	FT	/*tag= an	FT	CDS	complement (55135..56094)
FT	CDS	complement (34449..35210)	FT	FT	/*tag= bo
FT	FT	/*tag= ao	FT	FT	/product= "EvsC"
FT	FT	/product= "EvrN"	FT	FT	complement (56100..56103)
FT	RBS	complement (35219..35221)	FT	FT	/*tag= bp
FT	FT	/*tag= ap	FT	FT	complement (56184..56813)
FT	CDS	complement (35294..36238)	FT	FT	/*tag= bq
FT	FT	/*tag= aq	FT	FT	/product= "EvsC2"
FT	FT	/product= "EvrO"	FT	FT	56961..58709
FT	CDS	complement (36235..36963)	Query Match 12.7%; Score 51.6; DB 22; Length 109519;		
FT	FT	/*tag= ar	Best Local Similarity 50.8%; Pred. No. 0.2;		
FT	FT	/product= "EvrP"	Matches 123; Conservative 0; Mismatches 119; Indels 0; Gaps 0;		
FT	CDS	complement (36998..38026)	QY 162 gccgtgctgcctgcaggtggtgggagcaccagcgatgtcccggtccgacgagcgct 281		
FT	FT	/*tag= as	Db 40858 GGCACGCCGTTCCGACGACGCGGATCGCGACGTCGGCGCGGTCTCGACCCGAGGT 40799		
FT	FT	/product= "EvrQ"	QY 222 cctgatccgcctgcaggtggtgggagcaccagcgatgtcccggtccgacgagcgct 281		
FT	FT	complement (38072..38566)	Db 40798 GCTGGCCAGCTCAAGGCCGCGCGCTAAGCTGATCGACGTTCCGCGTCCGCGCGG 40739		
FT	CDS	/product= "EvrR"	QY 282 ggtggccttcctggaccgcacgcagcagatcgtgccgctgggagcagagcgctccctcgc 341		
FT	FT	/*tag= au	Db 40738 GGCACACGCCGACTACTTGGAACTACGAGTCCGAGGCGCGGACGCCGCTGTGTACCGGAT 40679		
FT	FT	/product= "EvrS"			
FT	CDS	complement (40216..40890)			
FT	FT	/*tag= av			
FT	FT	/product= "EvrT"			





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:25:01 ; Search time 165.56 Seconds  
(without alignments)  
603.847 Million cell updates/sec

Title: US-09-749-185-6

Perfect score: 407

Sequence: 1 atgagctctctgctcgga.....gaggagcagaacgcggctg 407

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*

- 1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2.6/ptodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2.6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54.4	13.4	3231	1	US-08-074-121-4
2	54.4	13.4	3231	5	PCT-US94-06447-4
3	51.8	12.7	1221	6	5212296-16
4	51.8	12.7	1879	6	5212296-5
5	49.2	12.1	4403765	4	US-09-103-840A-2
6	47.6	11.7	3353	1	US-08-453-695A-109
7	47.6	11.7	3353	2	US-08-268-161A-109
8	47.6	11.7	3353	2	US-08-453-702A-109
9	47.6	11.7	3353	4	US-09-099-639-109
10	47.6	11.7	3353	5	PCT-US95-08071-109
11	47.4	11.6	978	4	US-09-385-028-18
12	47.4	11.6	11604	4	US-09-385-028-13
13	47.4	11.6	15079	4	US-09-385-028-1
14	47.4	11.6	20235	1	US-07-642-734C-3
15	47.4	11.6	20235	3	US-08-439-009A-3
16	47	11.5	4411529	4	US-09-103-840A-1
17	46.8	11.5	36519	3	US-08-923-137-2
18	46.6	11.4	44377	2	US-08-804-227C-7
19	46.6	11.4	44377	2	US-08-804-198-1
20	46.2	11.4	2064	1	US-08-343-428-1
21	45.8	11.3	2824	2	US-09-010-928B-3
22	45.8	11.3	71989	4	US-09-443-501A-2
23	45.6	11.2	15872	4	US-09-105-537-1
24	45.4	11.2	1236	4	US-09-423-439-9
25	45.4	11.2	1926	4	US-09-423-439-25
26	45.4	11.2	1929	4	US-09-423-439-15
27	45.4	11.2	1974	4	US-09-423-439-59

28	45.4	11.2	1998	4	US-09-423-439-50	Sequence 50, Appl
29	45.4	11.2	2019	4	US-09-423-439-31	Sequence 31, Appl
30	45.4	11.2	2025	4	US-09-423-439-37	Sequence 37, Appl
31	45.4	11.2	2048	3	US-08-776-246-1	Sequence 1, Appl
32	45.4	11.2	2048	3	US-08-776-251-1	Sequence 1, Appl
33	45.4	11.2	3217	4	US-09-423-439-52	Sequence 52, Appl
34	45.4	11.2	4776	2	US-08-852-401-1	Sequence 1, Appl
35	45.2	11.1	1394	2	US-08-068-729-3	Sequence 3, Appl
36	45.2	11.1	1394	3	US-09-255-671-3	Sequence 3, Appl
37	44.4	10.9	1035	1	US-07-601-094-30	Sequence 30, Appl
38	44.4	10.9	1035	1	US-08-012-735-30	Sequence 30, Appl
39	44.4	10.9	1910	4	US-09-593-711A-3	Sequence 3, Appl
40	44.4	10.9	1914	1	US-07-601-094-1	Sequence 1, Appl
41	44.4	10.9	1914	1	US-08-012-735-1	Sequence 1, Appl
42	44.2	10.9	1611	2	US-08-551-211-4	Sequence 4, Appl
43	44.2	10.9	2202	2	US-08-551-211-5	Sequence 5, Appl
44	44.2	10.9	68750	3	US-09-335-402-1	Sequence 1, Appl
45	44.2	10.9	68750	4	US-09-568-109-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-074-121-4  
; Sequence 4, Application US/08074121  
; Patent No. 5767362  
; GENERAL INFORMATION:  
; APPLICANT: Best, Elaine  
; TITLE OF INVENTION: Methods and Compositions for Modulating  
; TITLE OF INVENTION: Lipid Content of Plant Tissues  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 2200 Sand Hill Road, Suite 100  
; CITY: Menlo Park  
; STATE: California  
; COUNTRY: US  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/074,121  
; APPLICATION NUMBER: US/08/074,121  
; FILING DATE: 08-JUN-1993  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rae-Venter, Barbara  
; REGISTRATION NUMBER: 32,750  
; REFERENCE/DOCKET NUMBER: 05938/043001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 854-5277  
; TELEFAX: (415) 854-0875  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3231 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 861..1328  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1349..2695  
; US-08-074-121-4



Db 372 cgaggtcgagaggtggtgcacggcttctctcgacgagatgctggcgccgcccagaccgc 431

QY 390 ggagcag 396

Db 432 cgacctg 438

#### RESULT 4

5212296-5  
; Patent No. 5212296  
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH  
; J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.  
; TEPPERMAN, JAMES M.  
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING  
; CYTOCHROMES  
; NUMBER OF SEQUENCES: 19  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/569,781  
; FILING DATE: 23-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 464,499  
; FILING DATE: 12-JAN-1990  
; APPLICATION NUMBER: 405,605  
; FILING DATE: 11-SEP-1989  
; SEQ ID NO: 5:  
; LENGTH: 1879  
5212296-5

Query Match 12.7%; Score 51.8; DB 6; Length 1879;

Best Local Similarity 50.6%; Pred. No. 0.015;

Matches 125; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 150 cgaggtcgccggtggtgcacgggagcgggagcgggagcgtccacatcgcccgcccgccgagagac 209.

Db 319 cggcgaccccggtgctctccacccgagcggagcagcaactcccgcccgccgacccg 378

QY 210 gttcgcgaggtctctgacgtccgctgagtggtggagcagcagcagcgtgttccgggtcgg 269

Db 379 ctctgagcgctcggtggagagccgagcggttcacgttcacgtcaagcgatcaaggcgtgccc 438

QY 270 cagcgccgctggtggtccttctgacccgacgagcaagatcgtgcgcgtggggcagga 329

Db 439 caccggcggtgagcgtacgtacgaggttcacgttcacgtcaagcgatcaaggcgtgccc 498

QY 330 ggttccctcgccgacttcgacccctgctcgacgagcgctggacgcatctctggccga 389

Db 499 cgaggtcgagaggtggtgcacggcttctctcgacgagatgctggcgccgcccagaccgc 558

QY 390 ggagcag 396

Db 559 cgacctg 565

#### RESULT 5

US-09-103-840A-2/C  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 12.1%; Score 49.2; DB 4; Length 4403765;

Best Local Similarity 46.7%; Pred. No. 0.042;

Matches 156; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 67 gactctacccggtgcggtgacgttctcatctgccccgagatgccccggtgacgtggcg 126

Db 2836407 GAAGCCCAAGGGCGGGGACGATCCTGTTGGCCCGGGGACCTGGCGCTGCGGATGGG 2836348

QY 127 ttcggcgagggtgtctcatcgacggaggttcgcgaggttcgccccggtgccccggtgacgtccac 186

Db 2836347 CTGCCGGTGTGGCGGTGGTGGCGAGTTCGGCGACGTCTCGGCGACGGCGTGCACACCTCG 2836288

QY 187 atcgccccccacccgagacgttcgcgaggttcgccccggtgccccggtgacgtggcg 246

Db 2836287 ATCCCGCCCGCGGCTGGCGCGCTGGGGCGGGCGCGCGGCGGCAAGGATTCACCGCTG 2836228

QY 247 gaccagcgatgttccggttcgacggcggtggtggtggttcttcggacccgacggac 306

Db 2836227 GCGCGGCGCTGGCCAAAGCTGGCGGTGGCGCGACGACGTGGCGGTTCATCTCCAAGCAC 2836168

QY 307 aagatctgcggtggggcaggagcgttccctcgccagcttcgacccctgctcgaagag 366

Db 2836167 GACACCTCGACGCTGCCAACGATCCACAGACGAGTGTGATGAACGCTGCGCGAC 2836108

QY 367 gcgtgacgcgtctcgtggcgagagcagacg 400

Db 2836107 GCCTGGCGGTTCGAGGGCGCGCGCTGTTTCG 2836074

#### RESULT 6

US-08-453-695A-109

; Sequence 109, Application US/08453695A

; Patent No. 5708143

; GENERAL INFORMATION:

; APPLICANT: Suzuki, Shintaro

; TITLE OF INVENTION: Protocadherin Materials and Methods

; NUMBER OF SEQUENCES: 115

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

; STREET: 233 South Wacker, 6300 Sears Tower

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/453,695A

; FILING DATE:

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 5708143and, Greta E.

; REGISTRATION NUMBER: 35,302

; REFERENCE/DOCKET NUMBER: 32658

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312/474-6300

; TELEFAX: 312/474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 109:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3353 base pairs

[illegible]

RESULT 7  
 US-08-268-161A-109  
 ; Sequence 109, Application US/08268161A  
 ; Patent No. 5798224  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 115  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
 ; ADDRESSEE: Borun  
 ; STREET: 233 South Wacker, 6300 Sears Tower  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/268,161A  
 ; FILING DATE: June 27, 1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Young J. Suh  
 ; REGISTRATION NUMBER: P-41,337  
 ; REFERENCE/DOCKET NUMBER: 27866/32149  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELE: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 109:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3353 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS

```

; LOCATION: 763..3123
US-08-268-161A-109

Query Match 11.7%; Score 47.6; DB 1; Length 3353;
Best Local Similarity 51.9%; Pred. No. 0.096;
Matches 107; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 126 gttcggcgcgagtggtcctcatcgacggaggttccgcgcgcgtgcggggagcggggacgtcca 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2301 GTTCGCTCTCCAGTCGCTGGACTACGAGGCCCTTCGAGGTTCCGGGTGGGCGC 2360

Qy 186 catcgccccgcgcacggagagcttcggcgaggtctctgataccgcctcgaggtggggag 245
    || || || || || || || || || || || || || || || || || || || || ||
Db 2361 CACAGACCGCGGCTTCGCGCGCTCAGCAGCGAGGCGCTGGTGGCAGTGCCTGGTCTGGA 2420

Qy 246 cgaccagcgatgttcgggttcgcgcgcgcgcctgtgacctctctggaccgcacgga 305
    || || || || || || || || || || || || || || || || || || || || ||
Db 2421 CGCAACGACAACTCGCCCTTCGTGCTGTACCCCGCTGCAGAACGGCTCCGGCCCTGCAC 2480

Qy 306 caagatcgtgcgcgtggggcaggagc 331
    || || || || || || || || || || || || || || || || || || || || ||
Db 2481 CGAGCTGCTGCCCGGCGGCGCGAGC 2506

RESULT 8
US-08-453-702A-109
; Sequence 109, Application US/08453702A
; Patent No. 5891706
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; ADDRESSEE: Borun
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patencin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,702A

```

CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5891706and, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 32657  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 109:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3353 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 763..3123  
 US-08-453-702A-109

Query Match 11.7%; Score 47.6; DB 2: Length 3353;  
 Best Local Similarity 51.9%; Pred. No. 0.096;



	Matches	107:	Conservative	0:	Mismatches	99:	Indels	0:	Gaps	0:
Qy	126	gttcggcgcgagtgctcatcgacgaggtccgcggcgtgcgggagcgaggagctcca	185							
Db	2301	gttcgctctccagtcgtggactacgagggccttcgaggttcgcggtggcgcc	2360							
Qy	186	catcgccgcgcgcagccgcgagacgttcggcgaggtcctgatcgcctcgaggtgggag	245							
Db	2361	cacagacggcggttcggcggtgcagcagggcgctggtgcgagtcgtgctcga	2420							
Qy	246	cgaccagcgcatgtcttcggggttcgcgcgcgcctggtgccttcctggaccgcacgga	305							
Db	2421	cgccaaagcacaactcggcccttcgtgctgtaccgcgtcgagacggcctccggccctgcac	2480							
Qy	306	caagatcgtgcgcgtggggcaggagc	331							
Db	2481	cgagctggtgcccgccggcgccgagc	2506							

RESULT 9  
US-09-099-639-109  
; Sequence 109, Application US/09099639  
; Patent No. 626237  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 115  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
; STREET: 233 South Wacker, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/099,639  
; FILING DATE: 18 JUN 1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/263,161  
; FILING DATE: 27 JUN 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greta E. No. 626237and  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/34703  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3353 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 763..3123  
US-09-099-639-109

```

Query Match      11.7% ; Score 47.6; DB 4; Length 3353;
Best Local Similarity 51.9% ; Pred. No. 0.096;
Matches 107; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 126 gtcgggggggagttgtctcatcgacggaggtccgcgcgtgcgggggacgggacgtcca 185
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Db	2301	GTTCGCTCTCCAGCTCGCTGGACTACAGAGCCCTTCAGAGCTTTCGAGTTCCCGCTGGGGGC	2360
Qy	186	catcgccgcgcgcacccgagacgttcggcgaggtcctgatccgcctgcaggtggggag	245
Db	2361	CACAGACCGCGCTTCGCGCGCTGAGCAGCAGAGCGCTGGTGCAGTCTCTGGTCTGGA	2420
Qy	246	cgaccaggcgatgtcccggttcggcagcgccgcttggtggccttcctggaccgcacgga	305
Db	2421	CGCCACGACAACCTCGCCCTTCGTGCTGTACCCGCTGCAGACGGCTCCGCGCCTGCAC	2480
Qy	306	caagatcgtagcgcctggggcaggagc	331
Db	2481	CGAGCTGTTGCCCGCGCGCGCAGC	2506

RESULT 10  
PCT-US95-08071-109  
Sequence 109, Application PC/TUS9508071  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSEE: Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/08071  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE: 23 DEC 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32149  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 109:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3353 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 763..3123  
PCT-US95-08071-109

	Query: Match	11.7%	Score 47.6;	DB 5;	Length 3353;
	Best local Similarity	51.9%;	Pred. No. 0.096;		
	Matches: 107;	Conservative	0;	Mismatches 99;	Indels 0; Gaps
Qy	126	gttcggcgaggagtgtctcatcagcaggaaggtccgcgcccgtgcggagcgaggagcgtcca	185		
Db	2301	gtttgcgctccagctcgctgacgttacgagcccttcagagctttcagagttccgcgtggcgccg	2360		

Db 628 GTCGACGACCGCGCGGATCGCCAAAGCTCAAGCGCTCTTACGGGGACGCGAAGACCCG 687  
Qy 253 gcgatgttccgggttcgacgagccgctgtggtcccttcctggaccgacacgaagatc 312  
Db 688: TTCCTCGGGTACGACCGCGAGCTGCTGGCGCGGAGGACCCCGGAGGAGGCCCGTC 747  
Qy 313 gtgcgctggggcaggagcgttccctcgcgacttcgacgcccctgctcagcagagcgctg 372  
Db 748: GCCATCTGTCCAGGCGCTCGACGATGTGACCGTGGGGTGAAGCTGTCGCCCGTGAC 807  
Qy 373: gaccgcatcctggccga 389  
Db 808: GTCCTCATCATCGACAA 824

RESULT 12  
US-09-385-028-13  
; Sequence 13, Application US/09385028  
; Patent No. 6232106  
; GENERAL INFORMATION:  
; APPLICANT: Susan E. Jensen  
; APPLICANT: Kwamena A Aidoo  
; APPLICANT: Ashish S. Paradkar  
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
; Patent No. 6232106  
; TITLE OF INVENTION: Acid Biosynthesis  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC  
; STREET: The Jenifer Bulding, 400 Seventh Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/385,028  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/790,462  
; FILING DATE: 29-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: D. Douglas Price  
; REGISTRATION NUMBER: 24,514  
; REFERENCE/DOCKET NUMBER: 1418/P57452US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 638-6666  
; TELEFAX: (202) 39305350  
; TELEX: RCA 248593 IDEA UR  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11604 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-385-028-13

Query Match 11.6%; Score 47.4; DB 4; Length 11604;  
Best Local Similarity 49.0%; Pred. No. 0.1;  
Matches 126; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
Qy 133 cgggagttgctcctacgacgaggttcgagccgctgcgggacgggacgtccacatcgcg 192  
Db 5146: CGGCGCGCTCTCTTCGACCGCAAGGTGCGCTGCTGGTGGAGCGCTTCCCGGCGGG 5205  
Qy 193 ccgcgcgacccgagacgttcgaggtcctgacccgctcaggtggggagcgaccag 252

Qy 186 cctgcgcccgcgacccgagagcttcgagaggtcctcctgacccgctcaggtggggag 245  
Db 2361 CACAGACCGCGCTTCGCCGCGCTGAGCAGCGAGCGCTGCTGGTGGTCTGGA 2420  
Qy 246 cgaccagcagatgttccgggttcgacgagccgctgtggtcccttcctggaccgacgga 305  
Db 2421 CGCCAAACGACAACTCGCCCTTCGTGCTGTACCCGCTCGAGAACGGCTCCGCGCCCTGCAC 2480  
Qy 306 caagatcgtcgcgctggggcaggaac 331  
Db 2481 CGAGCTGTGTGCCCCGCGGCGCGAC 2506

RESULT 11  
US-09-385-028-18  
; Sequence 18, Application US/09385028  
; Patent No. 6232106  
; GENERAL INFORMATION:  
; APPLICANT: Susan E. Jensen  
; APPLICANT: Kwamena A Aidoo  
; APPLICANT: Ashish S. Paradkar  
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
; Patent No. 6232106  
; TITLE OF INVENTION: Acid Biosynthesis  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC  
; STREET: The Jenifer Bulding, 400 Seventh Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/385,028  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/790,462  
; FILING DATE: 29-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: D. Douglas Price  
; REGISTRATION NUMBER: 24,514  
; REFERENCE/DOCKET NUMBER: 1418/P57452US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 638-6666  
; TELEFAX: (202) 39305350  
; TELEX: RCA 248593 IDEA UR  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 978 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-385-028-18

Query Match 11.6%; Score 47.4; DB 4; Length 978;  
Best Local Similarity 49.0%; Pred. No. 0.11;  
Matches 126; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
Qy 133 cgggagttgctcctacgacgaggttcgagccgctgcgggacgggacgtccacatcgcg 192  
Db 568 CGGCGCGCTCTCTTCGACCGCAAGGTGCGCTGCTGGTGGAGCGCTTCCCGGCGGG 627  
Qy 193 ccgcgcgacccgagacgttcgaggtcctgacccgctcaggtggggagcgaccag 252

Db 5206 GTCCAGACCGCGGCGGATCGCCAAAGCTCAAGCGCTCTACGGGGAGCGCGAAGACCGG 5265  
Qy 253 gcgatgttccgggtcgccagcgccctgtgtgcttcttcttgaccgacgagcaagatc 312  
Db 5266 TTCTCTGGGTACGACCGGAGCTCTGGCGCGGAGAGACCCCGGGACAAAGAGCGCGTC 5325  
Qy 313 gtgcgcgtggggcaggagcgttccctgcgcgaattcgcagccctgctcgcagagcgctg 372  
Db 5326 GCCCATCTGTCCAGGCGCTCGACGATGTGACCGTGGGGTGAAGCTCGTCCCGGTGAC 5385  
Qy 373 gaccgcatctctggccga 389  
Db 5386 GTCTCATCATCGACAA 5402  
RESULT 13  
US-09-385-028-1  
; Sequence 1, Application US/09385028  
; Patent No. 6232106  
; GENERAL INFORMATION:  
; APPLICANT: Susan E. Jensen  
; APPLICANT: Kwamena A Aidoo  
; APPLICANT: Ashish S. Paradkar  
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
; Patent No. 6232106  
; TITLE OF INVENTION: Acid Biosynthesis  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC  
; STREET: The Jenifer Bulding, 400 Seventh Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/385,028  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/790,462  
; FILING DATE: 29-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: D. Douglas Price  
; REGISTRATION NUMBER: 24,514  
; REFERENCE/DOCKET NUMBER: 1418/P57452US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 638-6666  
; TELEFAX: (202) 39305350  
; TELEX: RCA 248593 IDEA UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15079 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Streptomyces clavuligerus  
; US-09-385-028-1  
Query Match 11.68; Score 47.4; DB 4; Length 15079;  
Best Local Similarity 49.08; Pred. No. 0.1;  
Matches 126; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
Qy 133 cgggagtgctcatcgacgaggtccgcggtgcggtgagggagcggtccacatcgcg 192

Db 7178 CGGCGCGCTCTCTTCGACCGCAAGTGCCCTGCTGCTGGAGCTGGCTTCGCGGCGGG 7237  
Qy 193 ccgcgcacccgagagacgttcgagcgaggtcctctgacccctgcaggtgagagaccaa 252  
Db 7238 GTCCAGACCGCGGCGGATCGCCAAAGCTCAAGCGCTCTACGGGGAGCGCGAAGACCGG 7297  
Qy 253 gcgatgttccgggtcgccagcgccctgtgtgcttcttcttgaccgacgagcaagatc 312  
Db 7298 TTCTCTGGGTACGACCGGAGCTCTGCGCGCGGAGGACCCCGGGACAAAGAGCGCGTC 7357  
Qy 313 gtgcgcgtggggcaggagcgttccctgcgcgaattcgcagccctgctcgcagagcgctg 372  
Db 7358 GCCCATCTGTCCAGGCGCTCGACGATGTGACCGTGGGGTGAAGCTCGTCCCGGTGAC 7417  
Qy 373 gaccgcatctctggccga 389  
Db 7418 GTCTCATCATCGACAA 7434  
RESULT 14  
US-07-642-734C-3  
; Sequence 3, Application US/07642734C  
; Patent No. 5824513  
; GENERAL INFORMATION:  
; APPLICANT: Katz, L  
; APPLICANT: Donadio, S  
; APPLICANT: Mcalpine, J B  
; TITLE OF INVENTION: Recombinant DNA Method for Producing  
; TITLE OF INVENTION: Erythromycin Analogs  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Edward H. Gorman  
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
; STREET: Park Rd  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/642,734C  
; FILING DATE: 17-JAN-91  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Danckers, Andreas M  
; REGISTRATION NUMBER: 32652  
; REFERENCE/DOCKET NUMBER: 4952.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 708-937-9396  
; TELEFAX: 708-938-2623  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20235 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Saccharopolyspora erythraea  
; STRAIN: NRRL 238  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 19..10722  
; OTHER INFORMATION: /codon\_start= 19  
; OTHER INFORMATION: /function= "gene eryA"  
; OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for



```

Db 6764 TGGAC 6768
|||||
RESULT 15
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstein
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NRRL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722
; OTHER INFORMATION: /codon_start= 19
; OTHER INFORMATION: /function= "gene eryA"
; OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19..4470
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 97..1482
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1693..2670
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3406..3921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4171..4428
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4471..10722
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4471..5847
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6054..7026
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7165..9216
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: dehydratase and enoylreductase domains m"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9433..9984
; OTHER INFORMATION: /function= "approximate span
; OTHER INFORMATION: beta-ketoreductase of module 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10225..10483
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 4"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10723..20235
; OTHER INFORMATION: /codon_start= 10723
; OTHER INFORMATION: /function= "gene -eryA"
; OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
; OTHER INFORMATION: 6-deoxyerythronolide B formatio"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10723..15165
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10831..12174
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12379..13350
; OTHER INFORMATION: /function= "approximatr span of
; OTHER INFORMATION: acyltransferase domain of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14062..14610
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14857..15114
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 5"
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: 15166..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15172..16569
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16768..17721
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 18379..18921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
;
US-08-439-009A-3

```

```

Query Match      11.6%; Score 47.4; DB 3; Length 20235;
Best Local Similarity 47.2%; Pred. No. 0.1;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 71 cetaocgctgcggtgaccttctatctcccgagagatcccccggtgacctggcggttctg 130
   || || || || || || || || || || || || || || || || || || || || ||
Db 6464 CCGCGACGAGCTGGAGACCGTCTCGCGCTGGGACGCGCTCTCGCGGTGGCGCGG 6523

QY 131 ggcggagttgctcatcgagaggtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 190
   || || || || || || || || || || || || || || || || || || || || ||
Db 6524 TGAACGGGCGCTGGCACCGCTGCTGTCGGCGCGCGCGCGCGCGCGCGCGCGCG 6583

QY 191 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 250
   || || || || || || || || || || || || || || || || || || || || ||
Db 6584 CCGAGCGCGAGGCGCGGAGATGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6643

QY 251 aggcgatgttcccggttcggcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 310
   || || || || || || || || || || || || || || || || || || || || ||
Db 6644 CCGCGAGGTGGCGCGCATCGAGGACGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 6703

QY 311 tgcgtccgctgggcagagagcgttccctcgccgacttcgacccctgctcgacgagcgc 370
   || || || || || || || || || || || || || || || || || || || || ||
Db 6704 TCGGGGGCTCGGTGGCGCTGCACCTCCACGCTGACCGCGCGAGGTCAFCGACACCTCGCGCA 6763

QY 371 tggac 375
   || || || || || || || || || || || || || || || || || || || || ||
Db 6764 TGGAC 6768

```

Search completed: July 18, 2002, 11:34:02  
Job time: 11432 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 10:12:49 ; Search time 4821.96 seconds  
(without alignments)

1139.217 Million cell updates/sec

Title: US-09-749-185-6

Perfect score: 407

Sequence: 1 atgagcttcctcgtctcgga.....gaggagcagaacgcggctg 407

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	64.2	15.8	935	12	CNS006XK
2	62	15.2	925	12	CNS0091P
3	57.8	14.2	935	12	CNS006XK
C 4	57.2	14.1	925	12	CNS0091P
5	56.2	13.8	723	9	AU088650
6	53.2	13.1	558	10	C97336
C 7	53.2	13.1	1002	12	AG146122
8	53	13.0	932	12	CNS0072Q
9	51.8	12.7	633	9	AW564221
C 10	51.4	12.6	390	12	AQ846145
11	50.6	12.4	440	10	BI778054
12	50.4	12.4	474	12	AQ851336
13	50.4	12.4	616	12	CNS03M3T
14	49.8	12.2	397	9	BE125039
15	49.8	12.2	473	10	BE355969
16	49.8	12.2	652	9	BE194879
17	49.8	12.2	816	9	AL550433

18	49.8	12.2	1203	12	CNS015Y4
C 19	49.6	12.2	568	9	AW231567
20	49.4	12.1	440	12	AQ852321
C 21	49.2	12.1	374	9	AA234356
22	49.2	12.1	390	12	AQ846145
23	49.2	12.1	523	10	BI389979
24	49.2	12.1	563	10	BI392194
25	49.2	12.1	564	10	BI394352
26	49.2	12.1	619	9	AI981707
27	49.2	12.1	637	10	BM485736
28	49.2	12.1	655	10	BM426660
C 29	49.2	12.1	766	12	AG041031
C 30	49	12.0	525	9	AW076435
C 31	49	12.0	817	12	CNS0780L
32	49	12.0	832	9	AL519510
33	48.8	12.0	454	9	AV698167
34	48.8	12.0	502	12	AQ848096
35	48.8	12.0	528	10	BG274749
36	48.8	12.0	684	10	BI799714
37	48.8	12.0	704	9	AV688583
38	48.8	12.0	720	10	BF037954
39	48.8	12.0	979	9	AL543130
40	48.8	12.0	1039	9	AL579184
41	48.6	11.9	466	10	BM370573
42	48.4	11.9	327	9	AJ282193
C 43	48.4	11.9	371	12	AQ847078
44	48.4	11.9	1005	9	AL528164
45	48.2	11.8	435	9	AW746434

## ALIGNMENTS

CNS006XK 935 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC #  
BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
AL066051  
AL066051.1 GI:4945019  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
GSS.  
1 (bases 1 to 935)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
Location/Qualifiers  
1. .935  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"







```
/clone_lib="Rice mature leaf"
/tissue_type="mature leaf"
BASE COUNT      127 a      197 c      270 g      129 t
ORIGIN

Query Match      13.8%; Score 56.2; DB 9; Length 723;
Best Local Similarity 47.3%; Pred. No. 3.1;
Matches 169; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 37 attccggtgagctgcgttacgagacctgtatccctacgcgctgacggtgacgttccat 96
Db 23 ACTTCGTCGGAGTCGGAGTCGGAGTCGGAGTCGGAGTCGGAGTCGGAGTCGGAGTCGG 82

QY 97 ctgccgagagatccccggtgacctggcgttcggtggcggtggtgtctatcagcagagt 156
Db 83 CCGCGGCCCATGTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 142

QY 157 cgcgcccgtggtgggagggagctccacatcgccgcgcgcgcgcgcgcgcgcgcgcgc 216
Db 143 AAGAGGGGTGGCGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 202

QY 217 gaggctcctgacccgctgcaggtgggagcagcagcagcagcagcagcagcagcagcag 276
Db 203 GGGGTGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 262

QY 277 ccgctggtgacctcttcgacgcacgcacgcacgcacgcacgcacgcacgcacgcacgc 336
Db 263 CGCGGGGTGGCGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 322

QY 337 ctgcgcgaacttcgacccctctgcagcagcagcagcagcagcagcagcagcagcagcag 393
Db 323 GTGGCGGGGCACTACTACTCTGTACACGACGACGACGACGACGACGACGACGACGACG 379

RESULT 6
C97336      558 bp      mRNA      linear      EST 19-OCT-1998
LOCUS      C97336 Rice callus Oryza sativa cDNA clone C60023_11A, mRNA
DEFINITION      sequence.
ACCESSION      C97336
VERSION      C97336.1 GI:3760078
KEYWORDS      EST.
SOURCE      Oryza sativa.
ORGANISM      Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 558)
Sasaki,T. and Minobe,Y.
Rice cDNA from callus
Unpublished (1994)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL:http://rnp.dna.affrc.go.jp/
PROJECT "RGP",
Location/Qualifiers
1..558
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="C60023_11A"
/clone_lib="Rice callus"
/notes="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT      71 a      188 c      209 g      86 t      4 others
ORIGIN

Query Match      13.1%; Score 53.2; DB 12; Length 1002;
Best Local Similarity 49.1%; Pred. No. 9.4;
Matches 121; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 167 gcggggagcgggagcgtccacatcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 226
Db 294 GCGAGGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 353

QY 227 tcgcgctgcaggtgggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 286
Db 354 GCGAGTTCGTGCGGTGCGAGAGATGCGCGACCTGTTCTGTCGCCGCGCGCGCGCGCG 413

QY 287 ccttcctgacccgcacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 346
Db 414 GAGCTGCTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 473

QY 347 tcgacgcctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 403
Db 474 GCGGGCTGCTGCTGCGCATAGCGCTGCACCACTCNATCCGCGCGCGCGCGCGCGCGCG 530

RESULT 7
AG146122/c 1002 bp      DNA      linear      GSS 08-JAN-2002
LOCUS      AG146122 Pan troglodytes DNA, clone: RP43-007K01.TJ, genomic survey
DEFINITION      sequence.
ACCESSION      AG146122
VERSION      AG146122.1 GI:16675800
KEYWORDS      GSS; GSS (genome survey sequence);
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
Male BAC Library clone:RP43-007K01.TJ.
ORGANISM      Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library RPCI-43
Unpublished
2 (bases 1 to 1002)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asso Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the Rad process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
Location/Qualifiers
1..1002
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-007K01.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      57 a      404 c      455 g      25 t      61 others
ORIGIN

Query Match      13.1%; Score 53.2; DB 12; Length 1002;
Best Local Similarity 49.1%; Pred. No. 9.4;
```

[illegible]

RESULT	8	
CNS0072Q		
LOCUS		
DEFINITION		
ACCSSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES
source
Location/Qualifiers
1. .932
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RpC1-98"
/clone="BACR14B09"
/note="end : T7"

```

Query Match	13.0%	Score 53;	DB 12;	Length 932;
Best Local Similarity	32.1%;	Pred. No. 10;		
Matches 92;	Conservative 65;	Mismatches 130;	Indels 0;	Gaps 0;

Qy 75 cgcgctgcggctgacctttcatctgcccgagatgcccggtgacctggcggttcggcg 134  
 ||| | :: : | : ||| : | : ||| : | : | : | : | :  
 Db 628 cgcggsscgcsccgsgggcgcccgccgcsggsgsgsgcgccggcscgcgcs 687

[illegible]

## RESULT 9

AW564221	633 bp	mRNA	linear	EST 19-JUL-2000
LG1_289_A06.b1_A002	Light Grown 1 (LG1)	Sorghum bicolor	cdNA, mRNA	
sequence.				
AW564221				
AW564221.1	GI:7218099			
EST.				
	sorghum.			
	Sorghum bicolor			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC			
	clade; Panicoideae; Andropogoneae; Sorghum.			
	1 (bases 1 to 633)			
	Cordonnier-Pratt,M., Gingle,A., Marsala,C. and Pratt,L.H.			
	An EST database from Sorghum: light-grown seedlings			
	Unpublished (2000)			
	Contact: Cordonnier-Pratt MM			
	Department of Botany			
	The University of Georgia			
	Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA			
	Tel: 706 542 1860			
	Fax: 706 542 1805			
	Email: mmpratt@uga.edu			
	Sequences have been trimmed to exclude PolyA, vector and regions			
	below Phred quality 16. The threshold for highest quality sequence			
	is 20.			
	Seq primer: JEN REV			
	High quality sequence stop: 607			
	POLYA-No.			

FEATURES	Location/Qualifiers	BASE COUNT
source	1. .633	115 a 184 c 226 q 108 t
	/organism="Sorghum bicolor"	
	/db_xref="taxon:4558"	
	/clone_lib="Light Grown 1 (LGI)"	
	/note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda zap; Site.1: XhoI; Site.2: EcoRI"	
	: The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."	

Query Match	12.7%	Score 51.8;	DB 9;	Length 633;
Best Local Similarity	49.4%;	Pred. No. 16;		
Matches 134;	Conservative	0;	Mismatches 137;	Indels 0;
Gaps 0;				

[illegible][illegible]





AUTHORS	Roest-Crollius, H., Jaillon, O., Desilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizanes, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weltschbach, J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 616)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cps.fr/tetraodon">http://www.genoscope.cps.fr/tetraodon</a> .

FEATURES	source	Location/Qualifiers
1..616		
		/organism="Tetradodon nigroviridis"
		/db_xref="taxon:99883"
		/clone="O37G07"
		/clone_lib="G"
		/note="Genoscope sequence ID : COBG037AD04SP1-end : PUC-Orl"
BASE COUNT	135 a	185 c
ORIGIN	51 t	41 others

[illegible]

356 tgctgacgagcgctgagccgcatacctggccgaggagcagaa 398  
| | | | | | | | | | | | | | |  
517 ACCGCTCGGAGCTGNTCCACCCAGAGCCTGGCCGACGTGAAGAA 559

RESULT	14
BEL125039	
OCUS	
DEFINITION	
Bel125039	397 bp
DGL14_C09.b1_A002 Dark Grown 1 (DGL)	mRNA linear
sequence.	EST 19-JUL-2000
ACCESSION	
BEL125039	
VERSION	
BEL125039.1	GI:8547726
KEYWORDS	
EST.	

Sorghum, Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Lillipsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 397)  
Condonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
,L.H.  
TITLE An EST database from Sorghum: dark-grown seedlings  
JOURNAL Unpublished (2000)

CONTACT: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV  
High quality sequence stop: 382  
POLYA-NO.

FEATURES	source	Location/Qualifiers
		1. 397
		/organism="Sorghum bicolor"
		/db_xref="taxon:4558"
		/clone_lib="Dark Grown 1 (DGL)"
		/note="Organ: 5-day old dark-grown seedlings; Vector: Lambda zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT	62 a	174 c 99 g 62 t
ORIGIN		

	Query Match	12.2%	Score 49.8	DB 9	Length 397	
	Best Local Similarity	48.4%	Pred. No. 33			
	Matches 138	Conservative 0	Mismatches 147	Indels 0	Gaps 0	
Qy	122	gggqgttcggcgagttgtctcatcgcagcgagggtccgcggcgtgcyggggagcgggagc	181			
Db	103	GGAGGTTCCCGGAGGAGCTCGGCCCTGCTCGCGGACCTCGCGCTTCTCCACCTCAACTCA	162			
Qy	192	tcacatcgcccccgcgacccgagagagtttcggcgaggtctctgacccgctgcaggttg	241			
Db	163	ACCGCTTCGCGCGGCGCTGCGGAGTCTGCTCCCAAGCTGCGCTCTCTCCACGAGCTCG	222			
Qy	242	ggagcgacagggcgatgtttccgggttcggcacggcgcgctggtggccttcttgagccgca	301			
Db	223	ACGTGAGCAACAACCGCTTCACCGCGGGTTCCGCGAGCACATCTCTGCTGCCCAACA	282			
Qy	302	cggacacaagatctgcgctggggaggaagcgtttccctcgccgacttcgagccctctcg	361			
Db	283	TCAAGTACGTCGACCTCCCGGTTCAACAGGCTCTGCGGCGCGCTCCCGCGCGCTGTCG	342			
Qy	362	acgagggcgttgacccgcatactctggccgagagcaggaacgccggct	406			
Db	343	ACAAGCGCTCCGACGCCATCTCTCTCAACGAGCAACCACTTCGACT	387			

RESULT	15
BE355969	
LOCUS	473 bp mRNA linear EST 20-JUL-2000
DEFINITION	DG1_121.C03.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
ACCESSION	BE355969
VERSION	BE355969.1 GI:9297526
KEYWORDS	EST.
SOURCE	sorghum.
ORGANISM	Sorghum bicolor

ORGANISM	Sorghum bicolor Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE	1 (bases 1 to 473)
AUTHORS	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt L.H.
TITLE	An EST database from Sorghum: dark-grown seedlings
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV  
High quality sequence stop: 472

POLYA=NO.

FEATURES

source	Location/Qualifiers
	1..473
	/organism="Sorghum bicolor"
	/db_xref="taxon:4538"
	/clone_lib="Dark Grown 1 (DGI)"
	/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
BASE COUNT	75 a 198 c 128 g 72 t
ORIGIN	

Query Match 12.2%; Score 49.8; DB 10; Length 473;  
Best Local Similarity 48.4%; Pred. No. 33;  
Matches 138; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY	122	ggcgcttcggcgaggagtgctcatcgacgaggtccgcgcgcgtgcgggacgggacg	181
Db	99	GGAGTTCCCGGAGGAGCTGGGCTGCTCGCGACCTCGCGCTCCTCCACCTCAACTCCA	158
QY	182	tccacatcgcgcccgccgagacggttcggcgaggttcctgatccgcctgcagtg	241
Db	159	ACCGTTTCGCGCGGGGCTGCGGAGTCGCTCCCAAGCTGGGCTCCTCCACGAGCTCG	218
QY	242	ggagcgaccggcgatgttcgggtcgccacgcgcgcgtggtggccttcctggaccgca	301
Db	219	ACGTACGCAACAACCGCCTCACCGGGGGTTCCGCGAGCACATCCTCTGCTGCCCAACA	278
QY	302	cggacaagatcgtgcgcgtgggcgagagcgttccctccgcgactcgacgcctgctcg	361
Db	279	TCAAGTACGTGACCTCCGGTTCAAGAGCTCTCGGGCGCGTCCCGCGCGCTGTTCG	338
QY	362	acgagcgctggaccgcatacctggccgagagcagaacgcgcgt	406
Db	339	ACAAGCGCTCGAGCCATCTTCTCAACGACACACCACTTCGACT	383

Search completed: July 18, 2002, 10:12:53  
Job time: 9973 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:15:02 ; Search time 130.38 Seconds

(without alignments)  
115.010 Million cell updates/sec

Title: US-09-749-185-7

Perfect score: 701

Sequence: 1 MSFLVSEELSFRIPELVRYE.....FDALIDEALDRILAEQNAS 135

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	701	100.0	135	21	AAV44651
2	556	79.3	135	21	AAV44649
3	544	77.6	135	21	AAV44652
4	481	68.6	135	21	AAV44650
5	76	10.8	400	22	AAU56775
6	75.5	10.8	259	22	AA859590
7	74	10.6	736	22	AB859950
8	71.5	10.2	222	22	AAU55550
9	71	10.1	160	21	AA845480
10	71	10.1	540	22	AAU56042
11	70	10.0	257	21	AAG35780

12	70	10.0	258	21	AAG35779	Arabidopsis thalia
13	70	10.0	386	21	AAV81644	Streptococcus pneu
14	69	9.8	258	22	AAG92682	C glutamicum prote
15	69	9.8	258	22	AAB79361	Corynebacterium gl
16	69	9.8	398	21	AAG14865	Arabidopsis thalia
17	69	9.8	493	14	AAR38156	Acetobacter diquan
18	68.5	9.8	204	21	AAG11249	Arabidopsis thalia
19	68.5	9.8	253	21	AAG11248	Arabidopsis thalia
20	68.5	9.8	280	21	AAG11247	Arabidopsis thalia
21	68.5	9.8	473	22	AAB93677	Human protein sequ
22	68.5	9.8	890	20	AAV21669	Human hyd (h-Hyd)
23	68.5	9.8	897	22	ABB62660	Drosophila melanog
24	68.5	9.8	918	21	AAV84894	A human proliferat
25	68.5	9.8	2799	19	AAW81867	Human tumour suppr
26	68	9.7	411	22	AAU56274	Propionibacterium
27	68	9.7	413	22	AAB96344	Putative P. abyssi
28	68	9.7	623	22	AAB12571	Pseudomonas syring
29	68	9.7	1847	22	AAB31557	Pimaricin biosynth
30	67	9.6	798	22	AAB96715	Putative P. abyssi
31	66.5	9.5	1241	20	AAV42167	Human nephrin prot
32	66.5	9.5	1241	22	AAB47047	Human nephrin. Ho
33	66	9.4	162	22	AAG89844	C glutamicum prote
34	66	9.4	162	22	AAB79027	Corynebacterium gl
35	66	9.4	237	21	AA841315	Human ORF1079
36	66	9.4	309	22	AAG82257	S. epidermidis ope
37	66	9.4	507	22	AAG80035	Pseudomonas putida
38	66	9.4	507	22	AAG80043	Pseudomonas putida
39	66	9.4	592	22	AAU46702	Propionibacterium
40	66	9.4	6815	22	ABB66811	Drosophila melanog
41	65.5	9.3	201	22	ABB66188	Drosophila melanog
42	65.5	9.3	327	19	AAV85963	S. pneumoniae faeci
43	65.5	9.3	1333	22	AAU35343	Enterococcus faeci
44	65	9.3	286	22	AAB79061	Corynebacterium gl
45	65	9.3	302	22	AAU36303	Pseudomonas aerugi

#### ALIGNMENTS

RESULT 1

AAV44651  
ID AAV44651 standard; Protein; 135 AA.  
XX  
AC AAV44651;  
XX  
XX  
DT 18-APR-2000 (first entry)  
XX  
XX Streptomyces goldenensis SsgA protein.  
DE  
DE SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.  
XX  
OS Streptomyces goldenensis.  
XX  
PN WO200000613-A1.  
XX  
PD 06-JAN-2000.  
XX  
XX 25-JUN-1999; 99WO-NL00395.  
XX  
PR 26-JUN-1998; 98EP-0202148.  
XX  
PA (OYLE-) RIJKSUNIV LEIDEN.  
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZOE.  
XX  
PI Van Wezel GP, Kraal B, Luiten RGM;  
XX  
DR WPI; 2000-147269/13.  
DR N-PSDB; AAZ49730.  
XX



CC Filamentous bacteria can be transformed with ssgA gene-containing  
CC plasmid to enhance the production of secondary metabolites such as,  
CC antibiotics, antitumour agents, immunosuppressive agents,  
CC hypocholesterolaemic agents, enzyme inhibitors, antimigraine agents,  
CC herbicides, antiparasitic agents, ruminant growth promoters,  
CC bioinsecticides, receptor agonists and antagonists and biomass.

XX Sequence 135 AA;

Query Match 77.68; Score 544; DB 21; Length 135;  
Best Local Similarity 77.68; Pred. No. 1.5e-56;  
Matches 105; Conservative 12; Mismatches 18; Indels 0; Gaps 0;

Qy 1 MSFLVSELSFRIPIVELRYETCDPYAVRLTFHLPGDAPVTWAFGRELLIDGPRPCGDGD 60

Db 1 msflvseelsfripvelryetcdpyavrmthfhpdpvtwafgrellidgprpsgogd 60

Qy 61 VHAPADPETFGEVLIRLOVGSQDAMFRVGTAPLVAFIDRTDKIVPLGQERSLADFALL 120

Db 61 vhiaptdeglsvsirlqvadralfragapvlvafldrtksvpigqetlqdfeds1 120

Qy 121 DEALDRILAEQNAG 135

Db 121 eaalgkilaeeqnag 135

#### RESULT 4

AAU44650

ID AAY44650 standard; Protein: 135 AA.

XX AC AAY44650;

DT 18-APR-2000 (first entry)

XX Streptomyces albus G SsgA protein.

XX SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic; antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.

XX Streptomyces albus G.

XX WO200000613-A1.

XX 06-JAN-2000.

XX 25-JUN-1999; 99WO-NL00395.

XX 26-JUN-1998; 98EP-0202148.

XX (UYLE-) RIJKSUNIV LEIDEN.

XX (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.

XX Van Wezel GP, Kraal B, Luiten RGM;

XX WPI; 2000-147269/13.

XX N-PSDB; AAZ49729.

PT Reducing branching and enhancing fragmentation in filamentous  
PT microorganisms used to improve their liquid culturing properties

XX Disclosure; Fig 5; 60pp; English.

XX The present sequence is S. albus G SsgA protein. SsgA reduces branching  
CC and fragment separation and enhances fragmentation of mycelium in liquid  
CC culture resulting in lower viscosity of culture broths. Filamentous  
CC bacteria can be transformed with ssgA gene-containing plasmid to enhance  
CC the production of secondary metabolites such as, antibiotics, antitumour  
CC agents, immunosuppressive agents, hypocholesterolaemic agents, enzyme  
CC inhibitors, antimigraine agents, herbicides, antiparasitic agents,

CC ruminant growth promoters, bioinsecticides, receptor agonists and  
CC antagonists and biomass.

XX Sequence 135 AA;

Query Match 68.6%; Score 481; DB 21; Length 135;  
Best Local Similarity 71.1%; Pred. No. 4.5e-49;  
Matches, 96; Conservative 14; Mismatches 25; Indels 0; Gaps 0;

Qy 1 MSFLVSELSFRIPIVELRYETCDPYAVRLTFHLPGDAPVTWAFGRELLIDGPRPCGDGD 60

Db 1 msflvseelsfripvelryetcdpyavrltthfhpdpvtwvfgrellivegvladaagd 60

Qy 61 VHAPADPETFGEVLIRLOVGSQDAMFRVGTAPLVAFIDRTDKIVPLGQERSLADFALL 120

Db 61 vrvcpvgqtatrevnitlvgsqalfrvgkapllafldrtddqgisigserahadfnshl 120

Qy 121 DEALDRILAEQNAG 135

Db 121 ddalnrslaeeqsag 135

#### RESULT 5

AAU56775

ID AAU56775 standard; Protein: 400 AA.

XX AC AAU56775;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #17671.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59578.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
XX vaccinating against and diagnosing infections, especially useful for  
XX treating acne vulgaris

XX Claim 6; SEQ ID No 17970; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a



the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).  
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [www.wipo.int/pub/published/pct\\_sequences](http://www.wipo.int/pub/published/pct_sequences).

Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic polypeptides. The proteins and their associated DNA sequences are used in



XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0133456.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.

PR	29-SEP-1999;	99US-0156596.	PR	09-MAR-1999;	99US-0123548.
PR	04-OCT-1999;	99US-0157117.	PR	23-MAR-1999;	99US-0125788.
PR	05-OCT-1999;	99US-0157753.	PR	25-MAR-1999;	99US-0126264.
PR	06-OCT-1999;	99US-0157865.	PR	29-MAR-1999;	99US-0126785.
PR	07-OCT-1999;	99US-0158029.	PR	01-APR-1999;	99US-0127462.
PR	08-OCT-1999;	99US-0158232.	PR	06-APR-1999;	99US-0128234.
PR	12-OCT-1999;	99US-0158369.	PR	08-APR-1999;	99US-0128714.
PR	13-OCT-1999;	99US-0159293.	PR	16-APR-1999;	99US-0129845.
PR	13-OCT-1999;	99US-0159294.	PR	19-APR-1999;	99US-0130077.
PR	13-OCT-1999;	99US-0159295.	PR	21-APR-1999;	99US-0130449.
PR	14-OCT-1999;	99US-0159329.	PR	23-APR-1999;	99US-0130510.
PR	14-OCT-1999;	99US-0159330.	PR	23-APR-1999;	99US-0130891.
PR	14-OCT-1999;	99US-0159331.	PR	28-APR-1999;	99US-0131449.
PR	14-OCT-1999;	99US-0159637.	PR	30-APR-1999;	99US-0132048.
PR	18-OCT-1999;	99US-0159638.	PR	30-APR-1999;	99US-0132407.
PR	18-OCT-1999;	99US-0159584.	PR	04-MAY-1999;	99US-0132484.
PR	21-OCT-1999;	99US-0160741.	PR	05-MAY-1999;	99US-0132485.
PR	21-OCT-1999;	99US-0160767.	PR	06-MAY-1999;	99US-0132486.
PR	21-OCT-1999;	99US-0160768.	PR	06-MAY-1999;	99US-0132487.
PR	21-OCT-1999;	99US-0160770.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160814.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160815.	PR	14-MAY-1999;	99US-0134218.
PR	22-OCT-1999;	99US-0160980.	PR	14-MAY-1999;	99US-0134219.
PR	22-OCT-1999;	99US-0160981.	PR	14-MAY-1999;	99US-0134221.
PR	22-OCT-1999;	99US-0160989.	PR	14-MAY-1999;	99US-0134370.
PR	25-OCT-1999;	99US-0161404.	PR	18-MAY-1999;	99US-0134768.
PR	25-OCT-1999;	99US-0161405.	PR	19-MAY-1999;	99US-0134941.
PR	26-OCT-1999;	99US-0161406.	PR	20-MAY-1999;	99US-0135124.
PR	26-OCT-1999;	99US-0161359.	PR	21-MAY-1999;	99US-0135353.
PR	26-OCT-1999;	99US-0161360.	PR	24-MAY-1999;	99US-0135629.
PR	28-OCT-1999;	99US-0161361.	PR	25-MAY-1999;	99US-0136021.
PR	28-OCT-1999;	99US-0161920.	PR	27-MAY-1999;	99US-0136392.
PR	28-OCT-1999;	99US-0161992.	PR	28-MAY-1999;	99US-0136782.
PR	28-OCT-1999;	99US-0161993.	PR	01-JUN-1999;	99US-0137222.
PR	29-OCT-1999;	99US-0162142.	PR	03-JUN-1999;	99US-0137528.
Query Match 10.0%; Score 70; DB 21; Length 257;					
Best Local Similarity 30.7%; Pred. No. 6.4;					
Matches 31; Conservative 11; Mismatches 31; Indels 28; Gaps 5;					
QY	57	GDG-DVHIAPADPTFG-----EVLRLQVGSDOAMFRVGTAPLVAFLDRTDKIYPLGQ 109	PR	10-JUN-1999;	99US-0138094.
Db	38	gdgtsdsdppdpkpgdrrrqellari-----amqtskvrldfldrseyitkfa 91	PR	10-JUN-1999;	99US-0138540.
QY	110	ERSLADFDAL-----LDLALDRIL-----AERQNAQ 135	PR	10-JUN-1999;	99US-0138847.
Db	92	eeanaefdkvgdamkdleastrilenieskmgqfeesag 132	PR	14-JUN-1999;	99US-0139119.
RESULT 12					
AAG35779					
ID	AAG35779 standard; Protein; 258 AA.				
XX	AAG35779;				
AC					
XX					
DT	18-OCT-2000 (first entry)				
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 43757.				
XX					
XX					
KW	Protein identification; signal transduction pathway; metabolic pathway;				
KW	hybridisation assay; genetic mapping; gene expression control; promoter;				
KW	termination sequence.				
XX					
OS	Arabidopsis thaliana.				
XX					
PN	EP1033405-A2.				
XX					
PD	06-SEP-2000.				
XX					
PF	25-FEB-2000; 2000EP-0301439.				
XX					
XX	25-FEB-1999; 99US-0121825.				
PR	05-MAR-1999; 99US-0123180.				



PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144864.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156658.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 10.08; Score 70; DB 21; Length 258;

Best Local Similarity 30.7%; Pred. No. 6.4;

Matches 31; Conservative 11; Mismatches 31; Indels 28; Gaps 5;

QY 57 GDG--DVHIAPADPETFG-----EVLRLQVGSQAMFRVGTAPLVAFLDRDTKIVPLGQ 109

Db 39 gdgtsdsdsdpkpkpegdtrrqellari-----amigtksvrltdflderseyltkfa 92

QY 110 ERLSLADFAL-----LDEALDRIL-----AEEQNAG 135

Db 93 eanaefdkvgedamkdldeastrilenieskmaqafeesag 133

RESULT 13

AAV81644

ID AAV81644 standard; Protein; 386 AA.

AC AAV81644;

DT 24-MAY-2000 (first entry)

XX Streptococcus pneumoniae type 4 protein sequence #144.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;  
XX antibacterial; antinflammatory; meningitis; infection; diagnosis;  
XX pneumococcal disease.

OS Streptococcus pneumoniae.

PN W0200006737-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

XX 19-MAR-1999; 99US-0125164.

PA (MICR-) MICROBIAL TECHNIQS LTD.

XX

```

PI Gilbert CFG, Hansbro PM;
XX WPI; 2000-195300/17.
XX
XX New Streptococcal protein, useful as a vaccine, for diagnosis of
XX pneumococcal diseases and for screening agents capable of antagonizing
XX PT or inhibiting expression of the protein
XX
XX Claim 1; Page 87; 108pp; English.
XX
XX AA051501 to AA051679 represent specifically claimed protein sequences
XX isolated from Streptococcus pneumoniae. AA051501 to AA051590 represent
XX specifically claimed nucleotide sequences isolated from S. pneumoniae.
XX The sequences have antibacterial and antiinflammatory properties.
XX The protein sequences, and fragments of them, are useful as immunogens
XX and/or antigens. The nucleotide sequences can be used in vaccines and in
XX diagnostic assays. The proteins and nucleotides can be useful for the
XX detection and diagnosis of S. pneumoniae. The protein sequences are also
XX useful for screening an agent capable of antagonizing, inhibiting or
XX interfering with the function or expression of the proteins in which the
XX agent is useful for treatment or prophylaxis of S. pneumoniae infection
XX and meningitis. AA051591 to AA051614 represent primers used in the
XX exemplification of the present invention.
XX
XX Sequence 386 AA;
SQ

```

Query Match 10.0%; Score 70; DB 21; Length 386;  
Best Local Similarity 27.2%; Pred. No. 11;  
Matches 37; Conservative 18; Mismatches 39; Indels 42; Gaps 7;

```

QY 33 LPG----DAPVTWAFGRELIDGGPRPCGDDVHIAPAD-----PETFGEVLIRLQVGS 83
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 244 lpgtmiedylvefngkrfeavdgmgkpnepvervripedrlritlpeeg----klqkv 298
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 84 QAMFRVGTAPLVAF-----LDRTDKIVPLGQERSL-----ADFDAL 119
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 299 tqifrgvhyelaiydelgnwmihstrkal-vgeei gl dfepedihimrlhneteeefdar 357
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 120 LDEALDRILAEQNAG 135
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 358 lseeve---leeqag 370
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 14  
AAG92682  
ID AAG92682 standard; Protein; 258 AA.  
XX  
XX AAG92682;  
XX  
XX 26-SEP-2001 (first entry)  
XX  
XX C glutamicum protein fragment SEQ ID NO: 6436.  
XX  
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis.  
XX  
XX Corynebacterium glutamicum.  
XX  
XX EP1108790-A2.  
XX  
XX 20-JUN-2001.  
XX  
XX 18-DEC-2000; 2000EP-0127688.  
XX  
XX 16-DEC-1999; 99JP-0377484.  
XX  
XX 07-APR-2000; 2000JP-0159162.  
XX  
XX 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Nakagawa S, Mizouchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
XX Tateshi N, Senoh A, Ikeda M, Ozaki A;

```

XX WPI; 2001-376931/40.
XX N-PSDB; AAH67901.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 6436; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from Coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX
XX Sequence 258 AA;
SQ

```

Query Match 9.8%; Score 69; DB 22; Length 258;  
Best Local Similarity 26.1%; Pred. No. 8.5;  
Matches 24; Conservative 16; Mismatches 22; Indels 30; Gaps 5;

```

QY 48 LIDGGPRPCGCG-----DVHIA--PADPETFGEVLIRLQVGSQAMFRVGTAPLV 95
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6 mvtggaqgigrglseklaadgfdaiavadlpqeeqaetkllieaaggkavf-vg----- 59
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 96 AFLDRDTKIVPLGQERSLADFDALDEALDRI 127
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 --ldvtdk-----anfdsaideaaekl 79
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 15  
AAB79361  
ID AAB79361 standard; Protein; 258 AA.  
XX  
XX AAB79361;  
XX  
XX 30-APR-2001 (first entry)  
XX  
XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:238.  
XX  
XX Corynebacterium glutamicum; carbon metabolism and energy production;  
XX SMP protein; sugar metabolism and oxidative phosphorylation protein;  
XX fine chemical production; organic acid; proteinogenic amino acid;  
XX nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
XX nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
XX carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
XX diagnosis; Corynebacterium diphtheriae; evolutionary study.  
XX  
XX Corynebacterium glutamicum.  
XX  
XX WO200100844-A2.  
XX  
XX 04-JAN-2001.  
XX  
XX 23-JUN-2000; 2000WO-IB00943.  
XX  
XX 25-JUN-1999; 99US-0141031.  
XX  
XX 08-JUL-1999; 99DE-1031412.  
XX  
XX 08-JUL-1999; 99DE-1031413.  
XX  
XX 08-JUL-1999; 99DE-1031419.  
XX  
XX 08-JUL-1999; 99DE-1031420.  
XX  
XX 08-JUL-1999; 99DE-1031424.

PR	08-JUL-1999;	99DE-1031428.
PR	08-JUL-1999;	99DE-1031431.
PR	08-JUL-1999;	99DE-1031433.
PR	08-JUL-1999;	99DE-1031434.
PR	08-JUL-1999;	99DE-1031510.
PR	08-JUL-1999;	99DE-1031562.
PR	08-JUL-1999;	99DE-1031634.
PR	09-JUL-1999;	99DE-1032180.
PR	09-JUL-1999;	99DE-1032227.
PR	09-JUL-1999;	99DE-1032230.
PR	09-JUL-1999;	99US-0143208.
PR	14-JUL-1999;	99DE-1032924.
PR	14-JUL-1999;	99DE-1032973.
PR	14-JUL-1999;	99DE-1033005.
PR	27-AUG-1999;	99DE-1040765.
PR	31-AUG-1999;	99US-0151572.
PR	03-SEP-1999;	99DE-1042076.
PR	03-SEP-1999;	99DE-1042079.
PR	03-SEP-1999;	99DE-1042086.
PR	03-SEP-1999;	99DE-1042087.
PR	03-SEP-1999;	99DE-1042088.
PR	03-SEP-1999;	99DE-1042095.
PR	03-SEP-1999;	99DE-1042123.
PR	03-SEP-1999;	99DE-1042125.
XX	(BADI ) BASF AG.	
PA	Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;	
XX	WPI: 2001-061975/07.	
DR	N-PSDB; AAF71478.	
XX		
PT	New isolated Corynebacterium glutamicum nucleic acid encoding a sugar	
PT	metabolism and oxidative phosphorylation protein for production or	
PT	modulation of production of fine chemicals e.g. amino acids,	
PT	carbohydrates or enzymes -	
XX		
PS	Claim 20; Page 479-480; 1246pp; English.	
XX		
CC	AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar	
CC	metabolism and oxidative phosphorylation (SMP) proteins given in	
CC	AAB79243 to AAB 79633 which are involved in carbon metabolism and	
CC	energy production. The C. glutamicum SMP gene can be used in vectors	
CC	(II) for expression in host cells and production or modulation of	
CC	production of fine chemicals, such as, an organic acid, a proteinogenic	
CC	or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,	
CC	a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty	
CC	acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a	
CC	cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins	
CC	(III) encoded by them are used for diagnosing the presence or activity of	
CC	Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells	
CC	containing them are used to map genomes of organisms related to	
CC	C. glutamicum, identify and localise C. glutamicum sequences of interest,	
CC	in evolutionary studies, in determining SMP protein regions required	
CC	for function, in modulating SMP protein activity, in modulating the	
CC	metabolism of sugars, and in modulating high-energy molecule production	
CC	in a cell (i.e. ATP, NADPH).	
XX		
SQ	Sequence 258 AA;	
	Query Match 9.8%; Score 69; DB 22; Length 258;	
	Best Local Similarity 26.1%; Pred. No. 8.5;	
	Matches 24; Conservative 16; Mismatches 22; Indels 30; Gaps 5;	
Qy	48 LIDGGPCPGDG-----DVHIA--PADPETFGEVLIRLVGSDQAMFRVGTAPLV 95	
	:: :: ::     :  :	
Dd	6 mvtggagqgigisekllaadgfdiavadlpqeeaaetikiaagqkvf-vg----- 59	
	:: :: ::     :  :	
Qy	96 AFLDRTKIVPLQGERSLADFALLDREALRI 127	
	:: :: ::     :  :	
Dd	60 --lfdvdk-----afndsaidaeakl 79	
	:: :: ::     :  :	

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:16:07 ; Search time 51.31 Seconds  
(without alignments)  
64.265 Million cell updates/sec

Title: us-09-749-185-7

Perfect score: 701

Sequence: 1 MSFLVSEELSFRIPELVRYE.....FDALDELDRILAEQNAQ 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_AA.\*
- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75.5	10.8	259	1	US-08-015-986A-10
2	75.5	10.8	259	1	US-08-015-973-6
3	75.5	10.8	259	2	US-08-446-363-10
4	75.5	10.8	259	2	US-08-448-164-6
5	75.5	10.8	259	4	US-08-081-929-6
6	69	9.8	493	1	US-08-309-512-11
7	69	9.8	493	5	PCT-US92-08756A-11
8	66.5	9.5	1241	4	US-09-040-774-2
9	64.5	9.2	695	2	US-08-701-240-4
10	64.5	9.2	695	4	US-09-138-236-4
11	64.5	9.2	837	3	US-09-012-710-12
12	64.5	9.2	837	4	US-08-556-273-12
13	64	9.1	190	4	US-09-228-986-119
14	63	9.0	158	4	US-08-924-345-4
15	63	9.0	182	4	US-09-056-556-234
16	63	9.0	1100	1	US-08-357-598-11
17	63	9.0	1100	2	US-09-003-289-11
18	63	9.0	1100	5	PCT-US95-16435-11
19	63	9.0	5588	4	US-09-036-987A-6
20	63	9.0	5588	4	US-09-370-700-6
21	62.5	8.9	1044	2	US-08-777-405A-2
22	62.5	8.9	1044	2	US-08-977-871A-2
23	62.5	8.9	1044	2	US-09-225-951-2
24	61.5	8.8	142	4	US-09-228-986-122
25	61.5	8.8	153	4	US-09-228-986-118
26	61.5	8.8	306	2	US-08-484-905-75
27	61.5	8.8	306	3	US-08-481-985B-75

28	61.5	8.8	306	4	US-08-370-476-75	Sequence 75, Appl
29	61.5	8.8	694	2	US-08-701-240-2	Sequence 2, Appl
30	61.5	8.8	694	4	US-09-138-236-2	Sequence 2, Appl
31	61.5	8.8	3739	3	US-09-320-878-2	Sequence 2, Appl
32	61.5	8.8	3739	4	US-09-105-537-33	Sequence 33, Appl
33	61.5	8.8	11877	4	US-09-105-537-6	Sequence 6, Appl
34	61	8.7	106	2	US-08-715-554-3	Sequence 3, Appl
35	61	8.7	106	2	US-08-583-118-3	Sequence 3, Appl
36	61	8.7	236	1	US-08-266-570A-6	Sequence 6, Appl
37	61	8.7	253	5	PCT-US96-01314-53	Sequence 53, Appl
38	61	8.7	388	1	US-08-070-165F-4	Sequence 4, Appl
39	61	8.7	388	2	US-08-885-418-4	Sequence 4, Appl
40	61	8.7	1416	1	US-08-061-465-4	Sequence 4, Appl
41	60.5	8.6	290	2	US-08-484-905-80	Sequence 80, Appl
42	60.5	8.6	290	3	US-08-481-985B-80	Sequence 80, Appl
43	60.5	8.6	290	4	US-08-370-476-80	Sequence 80, Appl
44	60	8.6	178	4	US-09-056-556-236	Sequence 236, App
45	60	8.6	266	2	US-07-857-224B-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1  
US-08-015-986A-10  
; Sequence 10, Application US/08015986A  
; Patent No. 5532123  
; GENERAL INFORMATION:  
; APPLICANT: Schllessinger, Joseph  
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE  
; TITLE OF INVENTION: PHOSPHATASE-GAMMA  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/015,986A  
; FILING DATE: 10-FEB-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7683-028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 259 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-015-986A-10

Query Match 10.8%; Score 75.5; DB 1; Length 259;  
Best Local Similarity 28.2%; Pred. No. 0.23;  
Matches 31; Conservative 13; Mismatches 33; Indels 33; Gaps 7;  
QY 24 PYAVRLTFLHPGADPVYTW-----FGRELLIDGPPRCGPDVHIAPADPE--TFGEVL-- 75  
DB 86 PYRLR-QFHL-----HWGSSDDHSGSEHTVDGVKVA---AELHLVHNPKNYNTFKKALKQ 135



```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-081-929-6

Query Match      10.8%; Score 75.5; DB 4; Length 259;
Best Local Similarity 28.2%; Pred. NO. 0.23;
Matches 31; Conservative 13; Mismatches 33; Indels 33; Gaps 7;

QY 24 PYAVRLTFHPGDAPVTWA-----FGRELLIDGGPRPCGDGDVHLPADPE--TFGEVL-- 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86 PYRLR-QFHL-----HWGSSDDHGSHTYDGVKYA---AELHLVHNPKYNTFKEALKQ 135

QY 76 -----IKQVGSQAMFRVGTAPLAVFLDRTDKIVPLGQERSLADFD 117
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 RDGIAGVIGIEFKICHENGEPQI-----FLDALDKIKTKGKEAPFTKFD 178

RESULT 6
US-08-309-512-11
; Sequence 11, Application US/08309512
; Patent No. 5759828
; GENERAL INFORMATION:
; APPLICANT: Tal, Ronny
; APPLICANT: Benzinan, Moshe
; APPLICANT: Gelfand, David H.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Calhoun, Roger D.
; APPLICANT: Wong, Hing C.
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 2730 Sand Hill Road
; CITY: Menlo Park
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,512
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,218
; FILING DATE: 29-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bortner, Scott R.
; REGISTRATION NUMBER: 34,298
; REFERENCE/DOCKET NUMBER: 8145-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Acetobacter xylinum
US-08-309-512-11

```





```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,236
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/701,240
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-2032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 695 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-138-236-4

Query Match 9.2%; Score 64.5; DB 4; Length 695;
Best Local Similarity 21.3%; Pred. No. 19;
Matches 27; Conservative 23; Mismatches 40; Indels 37; Gaps

QY 34 PGDAPVTWAFGR-LLIDGGPR-----VAFLD---RTDKIVPLQGRSLADFALLDEALD 125
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 549 PVRSPITRSPARDSRRPSPRPSVHSTSSSDSHDSEENVPMNPNLSGE----- 600

QY 82 SDQAMFRV-----GTAPL-----VAFLD---RTDKIVPLQGRSLADFALLDEALD 125
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 601 -DPNLFASNSLDGGSPMNPKPGDKQVEYLDLDLSDGKSTPPRRKQKSSGGSSMADEVRD 659

QY 126 RLIAEEQ 132
   : : : |
Db 560 YVVVDQQ 666

RESULT 11
US-09-012-710-12
; Sequence 12, Application US/09012710
; Patent No. 608748
; GENERAL INFORMATION:
; APPLICANT: Vinkemeier, Uwe
; APPLICANT: Moarefi, Ismail
; APPLICANT: Darnell, Jr., James E.
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A
; TITLE OF INVENTION: STAT PROTEIN AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/012,710
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.

```

REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-194  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-012-710-12

Query Match 9.2%; Score 64.5; DB 3; Length 837;  
Best Local Similarity 25.3%; Pred. No. 25;  
Matches 41; Conservative 17; Mismatches 59; Indels 45; Gaps 9;  
QY 7 EELSFRIPV-ELRYETCDPYAVRLTFHLPDAPVTWAFGRELID---GGPRPCGD--- 58  
Db 126 EELKFTPLGRUHRVRETRLLRESLHL---GPKTGQVSLQNLIDPPLNGPGPSEDLPIT 182  
QY 59 -----GDVHIAPADPETTGEVLIRLQVSDQAMFRVGTAP---LVAFLDRTDKIV--- 105  
Db 183 LQGTVGDL-----TTQPLVLLRIQIWKRQOOLAGNGTTFEESLAGLQERGESLVEIYS 236  
QY 106 PLQGE-----RSLADFDALLDEAL-----DRILAEQ 132  
Db 237 QLHQEIGAASGELEPKTRASLSRLDEVLRTLTSSFLVEKQ 278

RESULT 12  
US-09-556-273-12  
Sequence 12, Application US/09556273  
Patent No. 6312887  
GENERAL INFORMATION:  
APPLICANT: Vinkemeier, Uwe  
APPLICANT: Moarefi, Ismail  
APPLICANT: Darnell, Jr., James E.  
APPLICANT: Kuriyan, John  
TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A  
TITLE OF INVENTION: STAT PROTEIN AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM: disk  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/556,273  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/012,710  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-194  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-556-273-12  
Query Match 9.2%; Score 64.5; DB 4; Length 837;  
Best Local Similarity 25.3%; Pred. No. 25;  
Matches 41; Conservative 17; Mismatches 59; Indels 45; Gaps 9;  
QY 7 EELSFRIPV-ELRYETCDPYAVRLTFHLPDAPVTWAFGRELID---GGPRPCGD--- 58  
Db 126 EELKFTPLGRUHRVRETRLLRESLHL---GPKTGQVSLQNLIDPPLNGPGPSEDLPIT 182  
QY 59 -----GDVHIAPADPETTGEVLIRLQVSDQAMFRVGTAP---LVAFLDRTDKIV--- 105  
Db 183 LQGTVGDL-----TTQPLVLLRIQIWKRQOOLAGNGTTFEESLAGLQERGESLVEIYS 236  
QY 106 PLQGE-----RSLADFDALLDEAL-----DRILAEQ 132  
Db 237 QLHQEIGAASGELEPKTRASLSRLDEVLRTLTSSFLVEKQ 278  
RESULT 13  
US-09-228-986-119  
Sequence 119, Application US/09228986  
Patent No. 6359198  
GENERAL INFORMATION:  
APPLICANT: Strabala, Timothy  
APPLICANT: Nieuwenhuizen, Niels  
TITLE OF INVENTION: Compositions Isolated from Plant Cells  
TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signalling  
FILE REFERENCE: 11000/1020  
CURRENT APPLICATION NUMBER: US/09/228,986  
CURRENT FILING DATE: 1999-01-12  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 119  
LENGTH: 190  
TYPE: PRT  
ORGANISM: Eucalyptus grandis  
US-09-228-986-119

Query Match 9.1%; Score 64; DB 4; Length 190;  
Best Local Similarity 23.5%; Pred. No. 3.9;  
Matches 19; Conservative 18; Mismatches 30; Indels 14; Gaps 4;  
QY 53 PRPCGSDGVHIAPADPETTGEVLIRLQVSDQAMFRVGTAPLVAFLD---RTDKIVPLGQ 109  
Db 32 PSCGSEDVHLVAVDDSLVDRKVI-----EHLKIKSSCKVTA-VDSGIRALQFLGLDE 83  
QY 110.ERSLADFALLDEALDRILAE 130  
Db 84 EKAAGDFNGL---KVDLIITD 101

RESULT 14  
US-08-924-345-4  
Sequence 4, Application US/08924345  
Patent No. 6224878  
GENERAL INFORMATION:  
APPLICANT: LEUNG-TACK Patricia  
APPLICANT: LEGASTELOIS Isabelle, Christine, Marie-Andree  
APPLICANT: AUDONNET Jean-Christophe, Francis  
APPLICANT: RIVIERE Michel, Emile, Albert  
TITLE OF INVENTION: Mutants and vaccines of the Infectious  
TITLE OF INVENTION: Bovine Rhinotracheitis virus



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:17:35 ; Search time 73.98 seconds  
(without alignments)  
175.345 Million cell updates/sec

Title: US-09-749-185-7

Perfect score: 701

Sequence: 1 MSFLVSEELSFRIPELVRYE.....FDALLDEALDRILAEQNAQ 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	589	84.0	136	2 T37179	probable regulator
2	215	30.7	142	2 T36147	probable regulator
3	168	24.0	138	2 T35247	probable regulator
4	131	18.7	142	2 T35319	probable regulator
5	84	12.0	1025	2 A53121	peroxisome assembl
6	81.5	11.6	291	2 H70678	hypothetical prote
7	78.5	11.2	260	2 A43641	carbonate dehydrat
8	78.5	11.2	892	2 T29420	probable transcrip
9	78	11.1	980	2 S71090	peroxisome biogene
10	77.5	11.1	407	2 D83460	cytochrome c-type
11	75.5	10.8	260	1 CRH03	carbonate dehydrat
12	75	10.7	343	2 I69009	MHC class I RT1.C-
13	74	10.6	175	2 B82380	purine-binding che
14	74	10.6	323	2 T09184	probable peroxidase
15	74	10.6	954	1 S68178	mixed-lineage prot
16	74	10.6	1387	2 T16511	hypothetical prote
17	73.5	10.5	259	2 A22612	carbonate dehydrat
18	73	10.4	559	1 S55383	peptidylprolyl iso
19	72.5	10.3	230	2 J14893	deoxyribonuclease
20	72.5	10.3	387	2 T44873	probable secreted
21	72.5	10.3	426	2 T08550	choline monooxygen
22	72.5	10.3	948	2 T07816	probable peroxin-6
23	72	10.3	218	2 A00089	probable basal-bod
24	71.5	10.2	757	2 G57472	penicillin-binding
25	71.5	10.2	757	2 AC2691	penicillin binding
26	71.5	10.2	996	2 G87687	hypothetical prote
27	71	10.1	346	1 H69789	probable alcohol d
28	70.5	10.1	289	2 H87608	conserved hypothet
29	70.5	10.1	600	2 H84176	medium-chain acyl-

30	70	10.0	385	2 F95161	hypothetical prote
31	70	10.0	385	2 E98027	hypothetical prote
32	70	10.0	457	2 T10360	late expression fa
33	69.5	9.9	163	2 G84282	molybdenum cofacto
34	69.5	9.9	310	2 C95950	conserved hypothet
35	69.5	9.9	400	1 SYZMCC	naringenin-chalcon
36	69.5	9.9	1165	2 A48667	peroxisomal assemb
37	69	9.8	322	2 G83766	quinone oxidoreduc
38	69	9.8	363	2 E89958	hypothetical prote
39	69	9.8	398	2 T05873	hypothetical prote
40	68.5	9.8	143	1 E70980	hypothetical prote
41	68.5	9.8	230	2 C87461	DnaA-related prote
42	68.5	9.8	662	2 S42826	probable ATPase
43	68.5	9.8	859	2 T01461	hypothetical prote
44	68.5	9.8	1417	2 H83132	probable sensor/re
45	68	9.7	237	1 A39017	modulation control

ALIGNMENTS

RESULT 1

T37179

probable regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T37179

R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: Z21598

A:Accession: T37179

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-136 <SEE>

A:Cross-references: EMBL:AL096823; PIDN:CAB46964.1; GSPDB:GN00070; SCOEDB:SCQ11.09

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: ssgA; SCOEDB:SCQ11.09

Query Match

Best Local Similarity 84.0%; Score 589; DB 2; Length 136;

Matches 115; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy	1	MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVTVAFGRLLIDGGPRCGDGD	60
Db	2	MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVTVAFGRLLIDGGPRCGDGD	61
Qy	61	VHIAPADPETFCEVLIRLQVSDQAMFRVGTAPLAVFLDRTDKIVPLGOERSLADFALL	120
Db	62	VRIAPVEPEPLAEVLIRLQVSDQALFRSSAAPLAVFLDRTDKIVPLGOEGALADFSLH	121

Qy

Db

Qy

Db

RESULT 2

T36147

probable regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000

C:Accession: T36147

R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: Z21598

A:Accession: T36147

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-142 <SEE>

A:Cross-references: EMBL:AL096852; PIDN:CAB51005.1; GSPDB:GN00070; SCOEDB:SCE19A.24

A:Experimental source: strain A3(2)

C:Genetics:



A:Molecule type: DNA  
A:Residues: 1-291 <COL>  
A:Cross-references: GB:Z81451; GB:AL123456; NID:g3261662; PIDN:CAB03770.1; PID:g1666159  
A:Experimental source: strain H37Rv  
C:Genetics:  
C:Superfamily: hypothetical protein MTH1814

Query Match 11.6%; Score 81.5; DB 2; Length 291;  
Best Local Similarity 27.4%; Pred. No. 1.3;  
Matches 37; Conservative 14; Mismatches 41; Indels 43; Gaps 6;

Qy 30 TFHLPDAPVTWAF-----GRELLIDGPRPCGDGDVHIADPADPTEFVILRLQV--GS 82  
Db 24 TGYLPDTATATAVFLADRLGRPLIVEG---PAGVGKTELARAVAGATGSLVRLQCYEGV 80  
Qy 83 DOA-----MFRVGTG-----PLVAFLDRTDKIVPLGQER 111  
Db 81 DEARALYEWNHAKQLRLIQAGSGDWEATKDVSEFLLQRLPLTAIRRTPTVLLIDET 140  
Qy 112 SLAD--FDALIDEAL 124  
Db 141 DKADIEIEGLLEVL 155

RESULT 7  
A43641  
carbonate dehydratase (EC 4.2.1.1) III - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 11-Feb-1993 #sequence\_revision 12-Mar-1993 #text\_change 22-Jun-1999  
C:Accession: A43641; A45400  
R:Tweedie, S.; Edwards, Y.  
Biochem. Genet. 27, 17-30, 1989  
A:Title: Mouse carbonic anhydrase III: nucleotide sequence and expression studies.  
A:Reference number: A43641; MUID:89227981  
A:Accession: A43641  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-260 <TWE>  
A:Cross-references: GB:M77796; NID:g192331; PIDN:AAA37355.1; PID:g309127  
A:Note: the authors translated the codon TAC for residue 114 as Leu  
R:Stanton, L.W.; Ponte, P.A.; Coleman, R.T.; Snyder, M.A.  
Mol. Endocrinol. 5, 860-866, 1991  
A:Title: Expression of CA III in rodent models of obesity.  
A:Reference number: A45400; MUID:92017893  
A:Accession: A45400  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 79-85,'G',87-93,'X',95-96 <STA>  
A:Note: sequence extracted from NCBI backbone (NCBIP:57846)  
C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
C:Keywords: carbon-oxygen lyase; hydro-lyase; zinc  
F:5-259/Domain: carbonic anhydrase homology <CAH>

Query Match 11.2%; Score 78.5; DB 2; Length 260;  
Best Local Similarity 28.2%; Pred. No. 2.3;  
Matches 31; Conservative 13; Mismatches 33; Indels 33; Gaps 7;

Qy 24 PYAVRLTFHLPDAPVTWA----FGRELLIDGPRPCGDGDVHIADPADP--ETFEVL-- 75  
Db 87 FYRLR-QFHL-----HWGSSDDHSEHTVDGVKVA---AELHLVHNPRYNTFGEALKQ 136  
Qy 76 -----IRLQVGDQAMFRVGTGAPLVAFLDRTDKIVPLGQERSLADFD 117  
Db 137 PDGIAVVGILLIKIGKEGFEQI-----LLDALDKIKTKGKAPPTTFD 179

RESULT 8  
T29420  
probable transcription regulator - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T29420  
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, July 1998  
A:Reference number: Z20619  
A:Accession: T29420  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-892 <PAR>  
A:Cross-references: EMBL:AL031155; PIDN:CAA20065.1  
C:Genetics:  
A:Note: SC3A7.02c  
C:Superfamily: regulatory protein malt

Query Match 11.2%; Score 78.5; DB 2; Length 892;  
Best Local Similarity 34.4%; Pred. No. 9.7;  
Matches 33; Conservative 8; Mismatches 32; Indels 23; Gaps 6;

Qy 12 RIPVE-LRYETCDPVAVRLTFHLPDAPVTWAFGRELLIDGPRPCG--DGDVHIADPADP 68  
Db 78 RVPTAWLIVETGD-----HPPG---VFWAYVFESL-----RVCGARAGDTPGAPADP 121  
Qy 69 ETFGEVLIRLQVGDQAMFRVGTGAPLVAFLDRTDKI 104  
Db 122 SGVGQRLL----ASLAALNALDSDPVVLVLDYDRM 153

RESULT 9  
S71090  
peroxisome biogenesis disorder group 4 protein PXAAAL - human  
C:Species: Homo sapiens (man)  
C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 02-Feb-2001  
C:Accession: S71090  
R:Yahraus, T.; Braverman, N.; Dotti, G.; Kalish, J.E.; Morrell, J.C.; Moser, H.W.; Val  
EMBO J. 15, 2914-2923, 1996  
A:Title: The peroxisome biogenesis disorder group 4 gene, PXAAAL, encodes a cytoplasm  
A:Reference number: S71090; MUID:96272151  
A:Accession: S71090  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-980 <YAH>  
A:Cross-references: EMBL:U56602; NID:g1354752; PIDN:AAC50655.1; PID:g1354753  
C:Superfamily: FtsH/SEC18/CDC48-type ATP-binding domain homology  
C:Keywords: ATP; nucleotide binding; P-loop  
F:470-477/Region: nucleotide-binding motif A (P-loop)  
F:719-928/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>  
F:744-751/Region: nucleotide-binding motif A (P-loop)

Query Match 11.1%; Score 78; DB 2; Length 980;  
Best Local Similarity 28.8%; Pred. No. 12;  
Matches 34; Conservative 14; Mismatches 48; Indels 22; Gaps 7;

Qy 30 TFHLPDAPVTWAFGRE---LLIDGPRPCGDGDVHIADPADPE---TF---GEVLIRL 78  
Db 720 TQPLEHPHLLSLGRLRSGLLLNGPP---GTGKTLAKAVATECSLTLFSLVKGPELINM 776  
Qy 79 QVGDQAMFRV-----GTAPLVAFLDRTDKIVPLGQERSLADFDALDEALDRILAE 130  
Db 777 YVQSEENVREVFARARAAAPCIIFDELDLAP-SRGRS-GDSGGVMDRVVSQLLAE 832

RESULT 10  
DB3460  
cytochrome c-type biogenesis protein PA1483 [imported] - Pseudomonas aeruginosa (stra  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: DB3460  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

Query Match	10.8%;	Score 75.5;	DB 1;	Length 260;
Best Local Similarity	28.2%;	Pred. No. 4.5;		

RESULT 12  
I69009  
MHC class I Rtl.C-type protein - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 21-Jan-2000  
C;Accession: I69009

Query Match	10.78;	Score 75;	DB 2;	Length 343;
Best Local Similarity	24.58;	Pred. No. 7;		
Matches 35;	Conservative	10;	Mismatches	36;
			Indels	62;
			Gaps	6;
Qy	10	SFRIPVELRYET---CDPYAVRLTFH-----	-----LFGDAPVTWAF-GRELLI	49
		:	:	
Db	191	SLRYLELRKETLLRSDDPKPVRLTHPREGDVTLCRMALGFYPADITLTWLQNGEDLTQ		250
		:	:	
Qy	50	D---GGPRPCGGDGHVHTAPADPETFGEVLRQLVGSDQAMFRVGTAPLVAFLDRTDKTYP		106

```

Qy 107 LQERS---LADFDALLDALDR 126
      .||:|:| | : : |
Db 275 LGKEQSYCTLVEHGLPEPLTQR 297

RESULT 13
B82380
purine-binding chemotaxis protein Chew VCA1094 [imported] - Vibrio cholerae (strain N
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82380
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:204056833
A:Accession: B82380
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-175 <HEI>

```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:31:54 ; Search time 45.63 Seconds  
(without alignments)  
114.555 Million cell updates/sec

Title: us-09-749-185-7  
Perfect score: 701  
Sequence: 1 MSFLVSEELSFRIPELVRYE.....FDALLDEALDRILAEQFNAG 135

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	84	12.0	1025	1	PEX6_YARLI
2	78.5	11.2	259	1	CAH3_MOUSE
3	78	11.1	978	1	PEX6_RAT
4	78	11.1	980	1	PEX6_HUMAN
5	76.5	10.9	259	1	CAH3_RAT
6	75.5	10.8	259	1	CAH3_HUMAN
7	74	10.6	256	1	GLO2_RHOC
8	74	10.6	954	1	M3KA_HUMAN
9	73.5	10.5	259	1	CAH3_HORSE
10	73	10.4	559	1	FKB7_WHEAT
11	72.5	10.3	230	1	DRNE_AERYH
12	72.5	10.3	948	1	PEX6_SCHPO
13	72.5	10.3	1165	1	PEX6_PICPA
14	70	10.0	419	1	IRF3_MOUSE
15	70	10.0	457	1	LEF4_NPVOP
16	69.5	9.9	400	1	CHS2_MAIZE
17	68.5	9.8	143	1	YW84_MYCTU
18	68.5	9.8	662	1	YWE1_SCHMA
19	68	9.7	237	1	NOLA_BRATA
20	68	9.7	318	1	ECHC_RHOSH
21	68	9.7	413	1	FTZ2_PIRAB
22	68	9.7	549	1	SRYA_DROPS
23	68	9.7	607	1	YJKO_YEAST
24	67.5	9.6	291	1	ISPE_TREPA
25	67.5	9.6	382	1	DXR_BACHD
26	67.5	9.6	1159	1	DP3A_VIBCH
27	67	9.6	414	1	FTZ2_PIRHO
28	67	9.6	450	1	HMMN_BRATA
29	67	9.6	484	1	HKX1_SCHPO
30	67	9.6	891	1	POL2_BAMMN
31	67	9.6	1101	1	GUNC_CELFI
32	66.5	9.5	1030	1	PEX6_YEAST
33	66.5	9.5	1091	1	SYI_TREPA

34 66 9.4 480 1 UCRL\_BOVIN P31800 bos taurus  
35 66 9.4 559 1 MERA\_ALCSP P94188 alicalignes  
36 66 9.4 1043 1 P11D\_MOUSE O35904 mus musculus  
37 65.5 9.3 237 1 DRNE\_AERYH O44064 aeromonas h  
38 65.5 9.3 256 1 YKE0\_YEAST P32860 saccharomyc  
39 65.5 9.3 429 1 GUNA\_BUTFI P22541 butyrivibri  
40 65.5 9.3 577 1 VP45\_YEAST P38932 saccharomyc  
41 65 9.3 310 1 HIS1\_SCHPO P40373 schizosacch  
42 65 9.3 454 1 Y856\_METH O26944 methanobact  
43 65 9.3 1067 1 TRIL\_STRCO O9de2 streptomyc  
44 64.5 9.2 837 1 STA6\_MOUSE P52633 mus musculus  
45 64.5 9.2 1886 1 GP21\_RAT P11654 rattus norv

ALIGNMENTS

RESULT 1  
PEX6\_YARLI  
ID PEX6\_YARLI STANDARD; PRT; 1025 AA.  
AC P36966;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Peroxisome biosynthesis protein PAY4 (Peroxin-6).  
GN PEX6 OR PAY4.  
OS Yarrowia lipolytica (Candida lipolytica).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Dipodascaceae; Yarrowia.  
OX NCBI\_TaxID:4952;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94103271; PubMed=8276851;  
RA Nutley W.M., Brade A.M., Eitzen G.A., Veenhuis M.  
RA Aitchison J.D., Szilard R.K., Glover J.R., Rachubinski R.A.;  
RT "PAY4, a gene required for peroxisome assembly in the yeast Yarrowia  
lipolytica, encodes a novel member of a family of putative ATPases.";  
J. Biol. Chem. 269:556-566(1994).  
RL CC - FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS.  
CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC - SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L23858; AAA16622.1; -  
DR InterPro: IPR003593; AAA.  
DR InterPro: IPR003960; AAA\_sub.  
DR InterPro: IPR003959; AAA\_subfam.  
DR Pfam: PF00004; AAA; 2.  
DR SMART: SM00382; AAA; 1.  
DR PROSITE: PS00674; AAA; 1.  
KW Peroxisome; ATP-binding.  
FT NP\_BIND 760 767 ATP (POTENTIAL).  
SQ SEQUENCE 1025 AA; 112258 MW; C41229CAE88AB7A6 CRC64;

Query Match 12.0%; Score 84; DB 1; Length 1025;  
Best Local Similarity 27.8%; Pred. No. 0.94;  
Matches 25; Conservative 12; Mismatches 37; Indels 16; Gaps 3;  
QY 55 PCGDDVHIAPADPTFF-----GEVLIRLQVGSQAMFR-----VGTAPLVAFLDR 100  
Db 761.PPGTGKTLAKAIATFTSLNFFSVKGPPELLNNYIGSEANVRVFKARDAPCVVFFDE 820  
QY 101 TDKIVPLQCRSLADFDALLDRILAE 130  
Db 821 LDSVAP--QRRNQDGGVMDRIVSQLAE 848

```
RESULT 2
ID CAH3_MOUSE STANDARD; PRT; 259 AA.
AC P16015;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbonic anhydrase III (EC 4.2.1.1) (Carbonate dehydratase III) (CA-III).
DE CA3 OR CAR3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89227981; PubMed=2496681;
RA Tweedie S., Edwards Y.;
RT "Mouse carbonic anhydrase III: nucleotide sequence and expression studies.";
RL Biochem. Genet. 27:17-30(1989).
CC -1- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) -> CO(2) + H(2)O.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27796; AAA37355.1; -
DR PIR; A43641.
DR HSSP; P00918; 1CIM.
DR SWISS-2DPAGE; P16015; MOUSE.
DR MGD; MGI:88270; Car3.
DR InterPro; IPR001148; Carb_anhydrase.
DR PRODom; PD000865; Carb_anhydrase; 1.
DR PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1.
KW Lyase; Zinc.
FT INIT_MET 0 0
FT METAL 93 93 ZINC (CATALYTIC).
FT METAL 95 95 ZINC (CATALYTIC).
FT METAL 118 118 ZINC (CATALYTIC).
SQ SEQUENCE 259 AA; 29397 MW; E35B5CDDCA4A54A93 CRC64;

Query Match 11.2%; Score 78.5; DB 1; Length 259;
Best Local Similarity 28.2%; Pred. No. 0.69;
Matches 31; Conservative 13; Mismatches 33; Indels 33; Gaps 7;

QY 24 PYAVRLTFHLPDAPYTW-----FGRELLIDGPRCGDGVHIAADP---EFFGEVL-- 75
DB 86 PYRLR-QFHL-----HWGSSDDHSGEHTVDGVKYA---AELHLVHNPNRYTFGEALQK 135
QY 76 -----IRLQVSGDQAMFRVGTAPLVAFLDRTDKIVPLGQERSLADFD 117
DB 136 PDGIYVVGILLKIGREKGEFQI-----LLDALDKIKTKGKEAPTFHD 178

RESULT 3
ID PEX6_RAT STANDARD; PRT; 978 AA.
AC P54777; O55097;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Peroxisome assembly factor-2 (PAF-2) (Peroxisomal-type ATPase 1) (Peroxin-6).
DE PEX6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=FISCHER 344; TISSUE=Liver;
RX MEDLINE=96083586; PubMed=7493019;
RA Tsukamoto T., Miura S., Nakai T., Yokota S., Shimozaawa N., Suzuki Y., Orii T., Fujiki Y., Sakai F., Bogaki A., Yasuno H., Osumi T.;
RT "Peroxisome assembly factor-2, a putative ATPase cloned by functional complementation on a peroxisome-deficient mammalian cell mutant.";
RL Nat. Genet. 11:395-401(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=SPRAGUE-DAWLEY;
RX Tsukamoto T., Hashiguchi N.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. REQUIRED FOR STABILITY OF THE PTS1 RECEPTOR.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN, PEROXISOME.
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63673; BAA09824.1; -
DR EMBL; D89660; BAA24931.1; -
DR EMBL; D89657; BAA24931.1; JOINED.
DR EMBL; D89658; BAA24931.1; JOINED.
DR EMBL; D89659; BAA24931.1; JOINED.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003960; AAA_sub.
DR InterPro; IPR003959; AAA_subfam.
DR Pfam; PF00004; AAA; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00674; AAA; 1.
KW Peroxisome; ATP-binding; Repeat.
FT NP_BIND 470 477 ATP (POTENTIAL).
FT NP_BIND 742 749 ATP (POTENTIAL).
FT MUTAGEN 476 476 K->A: NO LOSS OF FUNCTION.
FT MUTAGEN 748 748 K->A: LOSS OF FUNCTION.
FT CONFLICT 299 299 D -> G (IN REF. 2).
FT CONFLICT 333 333 V -> A (IN REF. 2).
FT CONFLICT 343 343 Q -> R (IN REF. 2).
FT CONFLICT 546 546 R -> C (IN REF. 2).
SQ SEQUENCE 978 AA; 104426 MW; F723193B7E95EA97 CRC64;

Query Match 11.1%; Score 78; DB 1; Length 978;
Best Local Similarity 28.8%; Pred. No. 3.7;
Matches 34; Conservative 14; Mismatches 48; Indels 22; Gaps 7;

QY 30 TFHLPGDAPVYTWAFGR-----LLIDGGPRCGDGVHIAADP-----TF---GEVLRL 78
DB 718 TIQPLEHPELLSLGLRRSGLLHGGP---GTGKTLKAVATECSTFLSVSGPELINM 774
QY 79 QVGSQDQAMFRV-----GTAPLVAFLDRTDKIVPLGQERSLADFDALLDALDRILAE 130
DB 775 YVQSEENVREYFARAAAAPCIIFDELDLSAP-SRGRS-GDSGGVMDRVVSQLLAE 830

RESULT 4
```





```

RX MEDLINE-98101484; PubMed-9440534;
RA Borghese R., Crimi M., Fava L., Melandri B.A.;
RT "The ATP synthase atpHAGDC (Fl) operon from Rhodobacter capsulatus.";
RL J. Bacteriol. 180:416-421(1998).
CC -!- FUNCTION: THIOLESTERASE THAT CATALYSES THE HYDROLYSIS OF S-D-
CC LACTOYL-GLUTATHIONE TO FORM GLUTATHIONE AND D-LACTIC ACID (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: (S)-(2-hydroxyacetyl)glutathione + H(2)O =
CC glutathione + a 2-hydroxy acid anion.
CC -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
CC -!- PATHWAY: GLYOXAL PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE GLYOXALASE II FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X99599; CAA67905.1; -.
DR HSSP; Q16775; LOH5.
DR InterPro; IPR001279; Beta_lactam_met.
DR Pfam; PF00753; lactamase_B; 1.
KW Hydrolase; Zinc.
FT METAL 56 ZINC 1 (BY SIMILARITY).
FT METAL 58 ZINC 1 (BY SIMILARITY).
FT METAL 60 ZINC 2 (BY SIMILARITY).
FT METAL 61 ZINC 2 (BY SIMILARITY).
FT METAL 61 ZINC 2 (BY SIMILARITY).
FT METAL 114 ZINC 1 (BY SIMILARITY).
FT METAL 133 ZINC 1 AND 2 (BY SIMILARITY).
FT METAL 171 ZINC 2 (BY SIMILARITY).
SQ SEQUENCE 256 AA; 27342 MW; B7E7E557A4B9A809 CRC64;

Query Match 10.6%; Score 74; DB 1; Length 256;
Best Local Similarity 25.4%; Pred. No. 2;
Matches 30; Conservative 16; Mismatches 48; Indels 24; Gaps 4;

QY 29 LTFHLP-----GDAPVTWAGRELLIDGGPRPCGDGDVHAPADPET-----70
DB 119 IAWLPFGAGLFGSLMSGCGR--LFEQTPAQMFDTLRLAALPPETRVCSGHEYTA 176
QY 71 FQEVILRLQVGSDQAMFRVGTAPLVAFLDRTDKIYPLGQERSLADF----DALLDEAL 124
DB 177 NGRFALSLEPGNGLHDMRDVRLALAGELPSLPTLGEERATNPFRLRADDALRAL 234

RESULT 8
M3KA_HUMAN STANDARD; PRT; 954 AA.
AC Q02779; Q12761; Q14871;
DT 01-FEB-1994 (Rel. 28, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mitogen-activated protein kinase kinase kinase 10 (EC 2.7.1.37)
DE (Mixed lineage kinase 2) (Protein kinase MST).
GN MAP3K10 OR MKK2 OR MST.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE-96128179; PubMed-8536694;
RA Dorow D.S., Devereux L., Tu G.F., Price G., Nicholl J.K.,
RA Sutherland G.R., Simpson R.J.;
RT "Complete nucleotide sequence, expression, and chromosomal
RL localisation of human mixed-lineage kinase 2.";
RL Eur. J. Biochem. 234:492-500(1995).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE-95249256; PubMed-7731697;
RA Katoh M., Hirai M., Sugimura T., Terada M.;
RT "Cloning and characterization of MST, a novel (putative)
RT serine/threonine kinase with SH3 domain.";
RL Oncogene 10:1447-1451(1995).
RN [3]
RP SEQUENCE OF 244-480 FROM N.A.
RC TISSUE=Colon epithelium;
RX MEDLINE-95238756; PubMed-8477742;
RA Dorow D.S., Devereux L., Dietzsch E., de Kretser T.;
RT "Identification of a new family of human epithelial protein kinases
RT containing two leucine/isoleucine-zipper domains.";
RL Eur. J. Biochem. 213:701-710(1993).
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND SKELETAL MUSCLE.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X90846; CAA62351.1; -.
DR EMBL; Z48615; CAA88531.1; -.
DR PIR; S32468; S32468.
DR HSSP; P29355; 1SEM.
DR MIM; 600137;
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR004040; STY_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRINTS; PR00109; TYRKINASE.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00221; STYKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50002; SH3; 1.
DR Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
KW ATP-binding; SH3 domain.
FT DOMAIN 2 5 POLY-GLU.
FT DOMAIN 16 81 SH3.
FT DOMAIN 98 360 PROTEIN KINASE.
FT BIND 104 112 ATP (BY SIMILARITY).
FT BIND 125 125 ATP (BY SIMILARITY).
FT ACT_SITE 222 222 BY SIMILARITY.
FT DOMAIN 384 405 LEUCINE-ZIPPER (BY SIMILARITY).
FT DOMAIN 419 440 LEUCINE-ZIPPER (BY SIMILARITY).
FT DOMAIN 449 463 ARG/LYS-RICH (BASIC).
FT CONFLICT 462 464 SRL -> AV (IN REF. 2).
FT CONFLICT 465 480 LKLRGGSHSLPSGF -> AQAGRRPHQPALWL (IN
FT REF. 3).
FT CONFLICT 471 471 G -> S (IN REF. 2).
FT CONFLICT 807 807 G -> R (IN REF. 2).
FT CONFLICT 818 818 V -> A (IN REF. 2).
SQ SEQUENCE 954 AA; 103623 MW; 538F4AAA559B0ABA CRC64;

Query Match 10.6%; Score 74; DB 1; Length 954;
Best Local Similarity 31.5%; Pred. No. 9.2;
Matches 28; Conservative 10; Mismatches 29; Indels 22; Gaps 5;

```

AC	01-OCT-1994 (Rel. 30, Created)	
AD	P39638;	
AE	01-OCT-1994 (Rel. 30, Last sequence update)	
AF	01-OCT-1994 (Rel. 30, Last sequence update)	
AG	15-JUL-1999 (Rel. 38, Last annotation update)	
AH	Extracellular deoxyribonuclease precursor (EC 3.1.21.-) (DNase)	
AI	DNS.	
AJ	Aeromonas hydrophila.	
AK	Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;	
AL	Aeromonas.	
AM	NCBI_TaxID=644;	
AN	[1]	
AO	SEQUENCE FROM N.A.	
AP		



```

DR Pfam; PF00004; AAA; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00674; AAA; FALSE_NEG.
KW Peroxisome; ATP-binding.
FT NP_BIND 695 702 ATP (POTENTIAL).
SQ SEQUENCE 948 AA; 106505 MW; F41DFD7DE6D391B5 CRC64;

Query Match 10.3%; Score 72.5; DB 1; Length 948;
Best Local Similarity 25.4%; Pred. No. 13;
Matches: 31; Conservative 19; Mismatches 49; Indels 23; Gaps 5;

Qy 27 VRLFHLPDGPVTFWAFGR-----LLIDGPRPCGDGVHIAPADPTF-----GEV 74
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 667 LRDTLQLPLQPELFPSOGLKPRSGVLLYGP--GTGKTLAKAVATSLSEFVSIKQPE 723
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Qy 75 LIRLQVSDQAMFR-----VGTAPLVAFLDRTKIVPLGOERSIADFDALLDEALDRIL 128
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 724 LLNMYGSEANRVNFEKARNSSPCVFFDELDSIAP--HRGNSSDGNVMDRVVSOLL 781
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Qy 129 AE 130
   :|
Db 782 AE 783

RESULT 13
PEX6_PICPA
ID PEX6_PICPA STANDARD; PRT; 1165 AA.
AC P33289;
DT 01-FEB-1994 (Rel. 28, Created)
DD 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Peroxisome biosynthesis protein PAS5 (peroxin-6).
GN PEX6 OR PAS5.
OS Pichia pastoris (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4922;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 11430;
RX MEDLINE=94043437; PubMed=8227124;
RA Spong A.P., Subramanani S.;
RT "Cloning and characterization of PAS5: a gene required for peroxisome
   biogenesis in the methylotrophic yeast Pichia pastoris.";
RL J. Cell Biol. 123:535-548(1993)
CC -!- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. NECESSARY FOR
   MATRIX PEROXISOMAL PROTEINS IMPORT. ESSENTIAL FOR GROWTH ON OLEIC
   ACID AND METHANOL AS SOLE CARBON SOURCE.
CC -!- SUBCELLULAR LOCATION: IT MAY BE ASSOCIATED WITH THE CYTOSKELETON
   OR IT MAY BE CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation in
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (see http://www.isb-sib.ch/announcement/
   or send an email to license@isb-sib.ch).
-----
CC EMBL; 222556; CAA80278.1; -.
CC InterPro; IPR003593; AAA.
CC InterPro; IPR003960; AAA_sub.
CC Pfam; PF00004; AAA; 2.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS00674; AAA; 1.
KW Peroxisome; ATP-binding.
FT NP_BIND 859 866
SQ SEQUENCE 1165 AA; 129137 MW; DEAI92CID5CA46EA CRC64;

```

```
Query Match 10.3%; Score 72.5; DB 1; Length 1165;
Best Local Similarity 23.5%; Pred. No. 16;
Matches 28; Conservative 19; Mismatches 49; Indels 23; Gaps 5;

QY 30 TTHLPDGPVWAGRE-----LLIDGPRCGDGVHIAADPTTF-----GEVLIR 77
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 834 TIDPMKHPELFSNGIKRSGILFYGP---GTGKTLAKAIATNFALNFVSKGPELLN 890
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 78 LQVGSQAMFRV-----GTAPLVAFLDRTDKIVPLGQERSLADFDALLDEALDRILAE 130
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 891 MYIGSEANVRKVFQARDAKPCVFFDELDSVAP--KRGNGDSEGVMDRIYSQLLAE 947
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 14
IRF3 MOUSE STANDARD; PRT; 419 AA.
AC P70671;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interferon regulatory factor 3 (IRF-3).
GN IRF3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BALB/C; TISSUE=Liver;
RC Hakem R., Grossman A., Antonio L., Suggs S., Mak T.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY ACTIVATE TRANSCRIPTION BY COMPLEX FORMATION WITH
CC OTHER TRANSCRIPTION FACTORS, POSSIBLY MEMBERS OF THE STAT
CC FAMILY. BINDS SPECIFICALLY TO THE IFN-STIMULATED RESPONSE ELEMENT
CC (ISRE) BUT NOT TO THE IRF-1 BINDING SITE PRD-1 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE IRF FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; U75839; AAB36924.1; -
EMBL; U75840; AAB36925.1; -
DR HSP; P23906; 2IRF.
DR MGD; MGI:1859179; Irf3.
DR InterPro; IPR001346; IRF.
DR Pfam; PF00605; IRF.1.
DR PRINTS; PR00267; INTERREGCT.
DR PRODOM; PD002355; IRF.1.
DR SMART; SM00348; IRF.1.
DR PROSITE; PS00601; IRF.1.
KW Transcription regulation; DNA-binding; Nuclear protein; Activator.
FT DNA_BIND 7 107 TRYPTOPHAN PENTAD REPEAT.
SQ SEQUENCE 419 AA; 46852 MW; 1FF67C4E0FC7F027 CRC64;

Query Match 10.0%; Score 70; DB 1; Length 419;
Best Local Similarity 32.5%; Pred. No. 9;
Matches 25; Conservative 7; Mismatches 19; Indels 26; Gaps 4;

QY 25 YAVRLTHLPDGPVWAGRELLIDGPRPCGDGVHIAADPTTFGEVLIRLQVGSQ 84
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 275 WAQRL-----CHSHAFWALGELLPSDSGRGP--DGEVH-----KDKDG 310
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 85 AMF--RVGTAPLVAFLD 99
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 311 AVFDRFFVADLIAFME 327
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

```
RESULT 15
LEF4 NPVOP STANDARD; PRT; 457 AA.
AC Q10340;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Late expression factor 4.
GN LEF-4.
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpmNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
CC -1- FUNCTION: REQUIRED FOR LATE AND VERY LATE GENE EXPRESSION (BY
CC SIMILARITY).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; U75930; AAC59090.1; -
DR Transcription regulation.
SQ SEQUENCE 457 AA; 51151 MW; E21E06E50BAC8390 CRC64;

Query Match 10.0%; Score 70; DB 1; Length 457;
Best Local Similarity 35.2%; Pred. No. 9;
Matches 25; Conservative 6; Mismatches 22; Indels 18; Gaps 3;

QY 78 LQVGSQAMFRV-----GTAPLVAFLDRTDKIVPLGQERSLAD-----FDALLD 121
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 162 LQLGSDAVLARVRLELEFECAAPAAASLDACELV--QMETLADHINIAPCLPYTTLLD 219
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 122 EALDRILAEQ 132
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 220 SATPRRFTREQ 230

Search completed: July 18, 2002, 14:31:56
Job time: 945 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:30:55 ; Search time 140.15 Seconds  
(without alignments)  
166.638 Million cell updates/sec

Title: US-09-749-185-7  
Perfect score: 701  
Sequence: 1 MSFLVSELSFRIPVELRYE.....FDALLDEALDRILAEQNAG 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	701	100.0	135	Q9F9B7	Q9F9B7 streptomyc
2	589	84.0	136	Q9X9D2	Q9X9D2 streptomyc
3	556	79.3	136	P95753	P95753 streptomyc
4	544	77.6	145	Q9F9B5	Q9F9B5 streptover
5	481	68.6	135	Q9F9B6	Q9F9B6 streptomyc
6	234	33.4	159	Q9L268	Q9L268 streptomyc
7	215	30.7	142	Q9S2F7	Q9S2F7 streptomyc
8	168	24.0	138	Q9X7M8	Q9X7M8 streptomyc
9	155	22.1	156	Q9FC07	Q9FC07 streptomyc
10	131	18.7	142	Q9X7R1	Q9X7R1 streptomyc
11	91.5	13.1	126	Q9RKC9	Q9RKC9 streptomyc
12	85	12.1	1459	Q9HG03	Q9HG03 penicillium
13	82.5	11.8	351	O52572	O52572 amycolatops
14	81.5	11.6	291	P71922	P71922 mycobacteri
15	79.5	11.3	1388	Q9C1E9	Q9C1E9 glomerella
16	78.5	11.2	260	Q9ERN8	Q9ERN8 mus musculu

17	78.5	11.2	261	11	Q9D1D7	Q9D1D7 mus musculu
18	78.5	11.2	892	2	O86603	O86603 streptomyc
19	78	11.1	981	11	Q99LC9	Q99LC9 mus musculu
20	77.5	11.1	407	16	Q913M9	Q913M9 pseudomonas
21	77	11.0	309	2	Q9AFC5	Q9AFC5 azospirillu
22	76.5	10.9	527	10	Q94K10	Q94K10 avicennia m
23	75	10.7	343	7	Q31271	Q31271 rattus norv
24	74	10.6	107	2	O87043	O87043 rattus chol
25	74	10.6	175	16	Q9KKL0	Q9KKL0 vibrio chol
26	74	10.6	323	10	P93548	P93548 spinacia ol
27	74	10.6	348	7	O77949	O77949 rattus norv
28	74	10.6	736	5	Q9W1Y0	Q9W1Y0 drosophila
29	74	10.6	1378	5	O21029	O21029 caenorhabdi
30	73.5	10.5	366	2	Q9K200	Q9K200 streptomyc
31	73	10.4	539	2	Q9L4V1	Q9L4V1 streptomyc
32	72.5	10.3	387	16	O69499	O69499 mycobacteri
33	72.5	10.3	407	2	Q931W2	Q931W2 streptomyc
34	72.5	10.3	426	10	Q9S2R0	Q9S2R0 arabidopsis
35	72	10.3	220	2	Q9EW7	Q9EW7 streptomyc
36	72	10.3	700	2	O34003	O34003 rhodobacter
37	71.5	10.2	996	16	Q9A2M0	Q9A2M0 caulobacter
38	71	10.1	346	16	O34788	O34788 bacillus su
39	71	10.1	389	16	Q92KZ1	Q92KZ1 rhizobium m
40	70.5	10.1	289	16	Q9A4D0	Q9A4D0 caulobacter
41	70.5	10.1	360	16	Q92PY5	Q92PY5 rhizobium m
42	70.5	10.1	445	2	Q9KH05	Q9KH05 bradyrhizob
43	70.5	10.1	459	2	O32384	O32384 synechococc
44	70.5	10.1	600	17	Q9HSM3	Q9HSM3 halobacteri
45	70.5	10.1	2179	12	Q91DM0	Q91DM0 petunia vel

## ALIGNMENTS

RESULT 1

Q9F9B7 PRELIMINARY; PRT; 135 AA.  
AC Q9F9B7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE SSGA.  
GN SSGA;  
OS Streptomyces goldeniensis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=121022;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA van Wezel G.P., Rousseau C., Kraal B.;  
RT "Cloning and sequencing of the Streptomyces goldeniensis ssgA gene."  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF195770; AAG28481.1; -;  
SQ SEQUENCE 135 AA; 14843 MW; 32006CC86BDE4ED6 CRC64;

Query Match 100.0%; Score 701; DB 2; Length 135;  
Best Local Similarity 100.0%; Pred. No. 3.3e-62;  
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFLVSELSFRIPVELRYETCDPYAVRLTHLPGDAPVTWAFGRELIDGGPRCGDGD 60

Db 1 MSFLVSELSFRIPVELRYETCDPYAVRLTHLPGDAPVTWAFGRELIDGGPRCGDGD 60

Qy 61 VHIAPADPETTGEVLIRLVQSGDQAMFRVGTAPLVAFILDRDVKIYPLGQERSLADFDALL 120

Db 61 VHIAPADPETTGEVLIRLVQSGDQAMFRVGTAPLVAFILDRDVKIYPLGQERSLADFDALL 120

Qy 121 DEALDRILAEQNAG 135

Db 121 DEALDRILAEQNAG 135

```
RESULT 2
Q9X902 ID Q9X902 PRELIMINARY; PRT; 136 AA.
AC Q9X902;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE REGULATOR.
GN SCQ11.09.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K., Harris D.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL096823; CAB46964.1; -.
SQ SEQUENCE 136 AA; 14920 MW; 4B67C1F1E0BEC88 CRC64;

Query Match 84.0%; Score 589; DB 2; Length 136;
Best Local Similarity 85.2%; Pred. No. 4.5e-51;
Matches 115; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 1 MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVPTWAFGRELIDGGPRPCGDG 60
Db 2 MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVPTWAFGRELIDGGPRPCGDG 61
QY 61 VHTAPADPETGFEVLIRLQVGSQAMFRVGTAPLVAFLDRTDKIVPLGQERSLADFALL 120
Db 62 VRTAPVEPELAEVLIRLQVGSQALFRSSAAPLVAFLDRTDKIVPLGQEGALADFSLH 121
QY 121 DEALDRILAEQNAQ 135
Db 122 DEALDRILAEQNAQ 136

Query Match 84.0%; Score 589; DB 2; Length 136;
Best Local Similarity 85.2%; Pred. No. 4.5e-51;
Matches 115; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 1 MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVPTWAFGRELIDGGPRPCGDG 60
Db 2 MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVPTWAFGRELIDGGPRPCGDG 61
QY 61 VHTAPADPETGFEVLIRLQVGSQAMFRVGTAPLVAFLDRTDKIVPLGQERSLADFALL 120
Db 62 VRTAPVEPELAEVLIRLQVGSQALFRSSAAPLVAFLDRTDKIVPLGQEGALADFSLH 121
QY 121 DEALDRILAEQNAQ 135
Db 122 DEALDRILAEQNAQ 136

RESULT 3
P95753 ID P95753 PRELIMINARY; PRT; 136 AA.
AC P95753;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SSGA.
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=B2682;
RA Shinichi K., Ensign J.;
RT "Cloning and characterization of a gene involved in sporulation and
RT cell division of Streptomyces griseus.";
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
[2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=B2682;
RA Kawamoto S., Ensign J.C.;
RT "Isolation of mutants of Streptomyces griseus that sporulate in
RT nutrient rich media.";
RL Nippon Hosenkin Gakkaiishi 9:124-135(1995).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=B2682;
RA Kawamoto S., Ensign J.C.;
RT "Cloning and characterization of a gene involved in regulation of
RT sporulation and cell division of Streptomyces griseus.";
RL Nippon Hosenkin Gakkaiishi 9:136-151(1995).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=B2682;
RX MEDLINE=97286526; PubMed=9141673;
RA Kawamoto S., Watanabe H., Hesketh A., Ensign J.C., Ochi K.;
RT "Expression analysis of a ssgA gene product associated with
RT sporulation and cell division in Streptomyces griseus.";
RL Microbiology 143:1077-1086(1997).
DR EMBL; D50051; BAA21558.1; -.
SQ SEQUENCE 136 AA; 14783 MW; C6A28A7823AD7C8B CRC64;

Query Match 79.3%; Score 556; DB 2; Length 136;
Best Local Similarity 78.5%; Pred. No. 8.6e-48;
Matches 106; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

QY 1 MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVPTWAFGRELIDGGPRPCGDG 60
Db 2 MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVPTWAFGRELIDGGPRPCGDG 61
QY 61 VHTAPADPETGFEVLIRLQVGSQAMFRVGTAPLVAFLDRTDKIVPLGQERSLADFALL 120
Db 62 VHTGPEPEGLGDVHRLQVGADRALFRAGTAPLVAFLDRTDKIVPLGQEHLDGDFDNL 121
QY 121 DEALDRILAEQNAQ 135
Db 122 EDALGRILAEQNAQ 136

RESULT 4
Q9F9B5 ID Q9F9B5 PRELIMINARY; PRT; 145 AA.
AC Q9F9B5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE SSGA.
GN SSGA.
OS Streptovorticillum netropsis (Streptovorticillum flavopersicus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=55404;
RN [1]
RN SEQUENCE FROM N.A.
RA van Wezel G.P., Rousseau C., Kraal B.;
RT "Cloning and sequencing of the Streptomyces netropsis ssgA gene.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF195772; AAG28483.1; -.
SQ SEQUENCE 145 AA; 15838 MW; 92AC0F30DF3D2620 CRC64;

Query Match 77.6%; Score 544; DB 2; Length 145;
Best Local Similarity 77.8%; Pred. No. 1.5e-46;
Matches 105; Conservative 12; Mismatches 18; Indels 0; Gaps 0;

QY 1 MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVPTWAFGRELIDGGPRPCGDG 60
Db 11 MSFLVSEELSFRIPELVRYETCDPYAVRLTFHLPDGPVPTWAFGRELIDGGPRPCGDG 70
QY 61 VHTAPADPETGFEVLIRLQVGSQAMFRVGTAPLVAFLDRTDKIVPLGQERSLADFALL 120
```



```
Db 77 VRWPSKTEGRSVVLVALSSPDGDLLEAPTPQVSAWLERTLRAVPPGTEGAQLGIDDL 136
Qy 121 DEALDR 126
   | | |
Db 137 AELLAR 142

RESULT 8
Q9X7M8 PRELIMINARY; PRT; 138 AA.
AC Q9X7M8;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PUTATIVE REGULATOR.
GN SC5F2A.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Kinashi H., Hopwood D.A.;
RL "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL049587; CAB40672.1;
SQ SEQUENCE 138 AA; 15261 MW; F531BA295144ABA5 CRC64;

Query Match 24.0%; Score 168; DB 2; Length 138;
Best Local Similarity 31.4%; Pred. No. 3.4e-09;
Matches 43; Conservative 17; Mismatches 61; Indels 16; Gaps 3;

Qy 1 MSFLVSELSFR-----IPVELRYETCDPYAVRLTFLPG---DAPVTWAFGREL 47
   | | : : : | | | : : : | | : : : | | : : : | | : : : |
Db 1 MSTVIEQSVEARLVAAPRMPSIPATLHYDRADPPAVRMTFPAPATLEGVEVCWTFREL 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 48 LIDGPGPCGDDVHIAPADPETGEVLRLQVGSQDQAMFRVGTAPLVAFLDRTDKIVPL 107
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 LIAGQFENGHGDVVR---PYAYDRTVIEFHAPGEGTAVIHVRSGELRRFLQAGELVPV 117
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 108 GQERSLADFADLLDEAL 124
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 118 GLEHLQLDLHDHDLAELM 134
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
Q9FC07 PRELIMINARY; PRT; 156 AA.
AC Q9FC07;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE PUTATIVE REGULATOR.
GN SC8A11.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Kinashi H., Hopwood D.A.;
RL "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL391041; CAC01575.1;
SQ SEQUENCE 156 AA; 17188 MW; 7181972EE7F3FB0 CRC64;

Query Match 22.1%; Score 155; DB 2; Length 156;
Best Local Similarity 35.3%; Pred. No. 7.7e-08;
Matches 41; Conservative 14; Mismatches 55; Indels 6; Gaps 2;

Qy 13 IPVELRYETCDPYAVRLTFLPGDAPVTWAFGRELIDGPPRCGDDVHI---APADPE 69
   | | : : : | | | : : : | | : : : | | : : : | | : : : |
Db 37 VPAEFYDPDHPPLLTIIRFAPEGAPPVTHVGRDILLHGLRRTSGLDGVQVWADPTDRE 96
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 70 TFGVEVLRLQVGSQDQAMFRVGTAPLVAFLDRTDKIVPLGQERSLADFADLLDEALD 125
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 T---AWLQVNAHGDAIESLPVPELEEWIDRTYLHVPACTESSRLGTDAFLSKLFD 149
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
Q9X7R1 PRELIMINARY; PRT; 142 AA.
AC Q9X7R1;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PUTATIVE REGULATOR.
GN SC5H1.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Kinashi H., Hopwood D.A.;
RL "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL049863; CAB42928.1;
SQ SEQUENCE 142 AA; 15741 MW; E24AA52C00AF40F1 CRC64;

Query Match 18.7%; Score 131; DB 2; Length 142;
```



DR InterPro: IPR001128; Cyt\_p450.  
DR Pfam: PF00067; p450; 1.  
DR PROSITE: PS00086; CYTOCHROME\_P450; UNKNOWN 1.  
KW Heme: Hypothetical protein: Monooxygenase: Oxidoreductase.  
SQ SEQUENCE 351 AA; 38512 MW; 2C390B56A60D5571 CRC64;

Query Match 11.8%; Score 82.5; DB 2; Length 351;  
Best Local Similarity 24.3%; Pred. No. 3.3;  
Matches 37; Conservative 23; Mismatches 53; Indels 39; Gaps 5;  
Qy 5 VSELSFRIP-----VELRYETCDPYAVRLTFLHLPDAPVT-----WAFGRELL 48  
Db 88 LSPEFAFPAPSLMICELGLVRYEDRAEFQQRASALLQWNPVAEAVKNADALRAFMQALV 147  
Qy 49 IDGPRCGD---GDVHTAPADPTFGEVLIRLVQSDQAMFRVGTAPLVAFLDRTDKIV 105  
Db 148 TDKRANPAGDIISLIHAGADPAL-----TDEELINIANLLIAGYDTTASML 196  
Qy 106 PLG-----QERSLADFALLDEALDRIL 128  
Db 197 GLGIFVLLQRPALATLRDPSRIADAVEGLL 228

RESULT 14  
P71922 PRELIMINARY; PRT; 291 AA.  
AC P71922:  
DT 01-FEB-1997 (Tremblrel. 02, Created)  
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE HYPOTHETICAL 31.9 KDA PROTEIN.  
GN RV2426C OR MTCY428.21.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
SEQUENCE FROM N.A.  
RY STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
DR EMBL: Z81451; CAB03770.1; -.  
DR TubercuList; RV2426C; -.  
DR InterPro: IPR003959; AAA\_subfam.  
DR Pfam: PF00004; AAA; 1.  
KW Hypothetical protein: Complete proteome.  
SQ SEQUENCE 291 AA; 31950 MW; FA611423E14B4DCF CRC64;

Query Match 11.6%; Score 81.5; DB 16; Length 291;  
Best Local Similarity 27.4%; Pred. No. 3.3;  
Matches 37; Conservative 14; Mismatches 41; Indels 43; Gaps 6;  
Qy 30 TFHLPDAPVTWAF-----GRELLIDGPRPCGDDVHIAPADPTFGEVLIRLVQV--GS 82  
Db 24 TGYLPOTATATAVELADRLCKPLLEV---PAGVGKTELARAVAQAATGSLVRLQCYEGV 80  
Qy 83 DOA-----MFRVGT-----PLVAFLDRTDKIVPLGQER 111  
Db 81 DEARALYENNHAKOILRIQAGSGDWKATKTVFSEELQRPILLTAIRRTPTVLLIDET 140  
Qy 112 SLAD--FDALLDEAL 124

Db 141 DKADIEIEGLLEVL 155  
RESULT 15  
Q9C1E9 PRELIMINARY; PRT; 1388 AA.  
AC Q9C1E9:  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PEX6; PROTEIN.  
GN PEX6.  
OS Glomerella lagenarium (Anthracnose fungus) (Colletotrichum  
OS lagenarium).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;  
OC mitosporic Phyllachoraceae; Colletotrichum.  
OX NCBI\_TaxID=5462;  
RN [1]  
SEQUENCE FROM N.A.  
RY MEDLINE=21380423; PubMed=11487704;  
RX Kimura A., Takano Y., Furusawa I., Okuno T.;  
RT "Peroxisomal Metabolic Function Is Required for Appressorium-Mediated  
RT Plant Infection by Colletotrichum lagenarium."  
RL Plant Cell 13:1945-1957(2001).  
DR EMBL: AF343063; AAK16738.1; -.  
DR InterPro: IPR003593; AAA.  
DR InterPro: IPR003960; AAA\_sub.  
DR InterPro: IPR003959; AAA\_subfam.  
DR Pfam: PF00004; AAA; 2.  
DR SMART: SM00382; AAA; 1.  
DR PROSITE: PS00674; AAA; 1.  
KW ATP-binding  
SQ SEQUENCE 1388 AA; 150062 MW; 05E63C1D011531EE CRC64;  
Query Match 11.3%; Score 79.5; DB 3; Length 1388;  
Best Local Similarity 25.2%; Pred. No. 34;  
Matches 31; Conservative 18; Mismatches 51; Indels 23; Gaps 5;  
Qy 26 AVRLTFHLPDAPVTWAFGR-----LLIDGPRPCGDDVHIAPADPTF-----GE 73  
Db 1005 AVTETIQPLERPELFAKMKRSGILFYGPP---GTGKTLAKAIATKYSLNFFSVKGP 1061  
Qy 74 VLIRLVQSDQAMFR-----VGTAPLVAFLDRTDKIVPLGQERSLADFALLDEALDRI 127  
Db 1062 ELLNMYIGSEANVRVRFQARDARPCVFFDELDSVAP--KRGNGDGGVMDRIVSQL 1119  
Qy 128 LAE 130  
Db 1120 LAE 1122  
Search completed: July 18, 2002, 14:30:57  
Job time: 946 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM, nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:22:25 ; Search time 4085.76 Seconds  
(without alignments)  
2084:582 Million cell up

Title:	US-09-749-185-8
perfect score:	407
Sequence:	1 atgagcttcctcgtctccga.....gaggagcagaacgcgcgctg 407

Scoring table:    IDENTITY\_NUC  
Gapop 10.0 ,    Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*  
2: gb\_hug:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_o:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
<p>8</p>						

RESULT	1
AX007224	
LOCUS	407 bp DNA linear PAT 06-SEP-2000
DEFINITION	Sequence 9 from Patent WO0000613.
ACCESSION	AX007224
VERSION	AX007224.1 GI:9995090
KEYWORDS	
SOURCE	Streptomyces netropsis.
ORGANISM	Streptomyces netropsis
REFERENCE	Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
AUTHORS	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
TITLE	1 (bases 1 to 407)
JOURNAL	Kraal, B., Luiten, R.G. and Van Wezel, G.P. Reducing branching and enhancing fragmentation in culturing filamentous microorganisms Patent: WO 0000613-A 9 06-JAN-2000; UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GILSBERTUS MARIA (NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS VAN (NL)
FEATURES	Location/Qualifiers
source	1..407 /organism="Streptomyces netropsis" /db_xref="taxon:55404"

gene	1. .405	CDS	/db_xref="taxon:55404"					
	/gene="ssgA"							
	1. .>405							
	/gene="ssgA"							
	/note="strain ATCC of Streptomyces netropsis"							
	/codon_start=1							
	/transl_table=11							
	/product="SsgA"		/product="SsgA"					
	/protein_id="CAC07388.1"							
	/db_xref="GI:9995091"							
	/translation="MSFLYSELSFKIPVELRYETRDYPVAVRMTFHLPCDAPVTFWAFG							
	RELLDGINRPSGDGVHIAPDPEGLSDVSRILQVGADRALFRAGAPPLVAFDLDRD							
	KSVPLGQEQTLGDFDSLEAALGKILAEQNAG"							
BASE COUNT	63 a 151 c 131 g 62 t							
ORIGIN	70 a 158 c 142 g 68 t							
Query Match 100.0%; Score 407; DB 6; Length 407;								
Best Local Similarity 100.0%; Pred. No. 2e-38;								
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;								
QY	1	atgagttctctgcgcgagagctctctcaagatcccaatcccaatcgaaactcgatacagag	60					
Db	1	ATGAGCTTCTCTGCTCCGAGGAGCTCTCTTCAAGATCCCACTCGAATCGCATACGAG	60					
QY	61	accgggatactacgcggtgcgagatgacctccacctcccgagagacgcgcctgtgacc	120					
Db	61	ACCGGGATCCCTACGGGTGCGGATGACCTTCCACCTCCCCGGAGACGCGCTGTGACC	120					
QY	121	tgggcttgcgcgagctgtctgcagggatacaaccgccgagcgagcgagcgagcgac	180					
Db	121	TGGGCTTTCGGCGGGAGTGTCTGCTCGACGGGATCAACGCCCGGACGCGCGAC	180					
QY	181	gtccacatgcgcgcgacccgagggcctctcgagcttccatcccgctccaggtg	240					
Db	181	GTCACATGCCCCCGACCCGAGGGCTGTGCGAGCTCTCCATCCGCTCCAGGTG	240					
QY	241	ggcgagacgcgcctcttcctgcagggcgcccgccgctgccttcctccgaccg	300					
Db	241	GGCGGACCGCGCCCTCTTCCTGTCAGGGCGCCCGCGCTGCTCCCTCCGACCGC	300					
QY	301	acggacaagtgcgtgcgcgcgaggaacagactctgggtgacttcgagagacagcctg	360					
Db	301	ACGGACAAGTGCCTGCGCTCGTTCAGGAACAGACTCTGGGTGACTTCGAGGACGCTG	360					
QY	361	gagccgcgctcggaagatcttcgcgagagcagaaacgcgcgctg	407					
Db	361	GAGCCGCGCTCGCAAGATCTTCGCGGAGGACGAGAACCCCGCTG	407					
RESULT	2							
AF195772	AF195772	438 bp	DNA linear BCT 01-NOV-2000					
LOCUS	Streptomyces netropsis SsgA (ssgA) gene, complete cds.							
DEFINITION								
ACCESSION	AF195772							
VERSION	AF195772.1	GI:11066162						
KEYWORDS	Streptomyces netropsis.							
SOURCE	Streptomyces netropsis.							
ORGANISM	Streptomyces netropsis.							
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces.							
AUTHORS	van Wezel, G.P., Rousseau, C. and Kraal, B.							
TITLE	Cloning and sequencing of the Streptomyces netropsis ssgA gene							
JOURNAL	Unpublished							
REFERENCE	2 (bases 1 to 438)							
AUTHORS	van Wezel, G.P.							
TITLE	Direct Submission							
JOURNAL	Submitted (18-OCT-1999) Biochemistry, University of Leiden, PO Box 9502, Leiden 2300 RA, Netherlands							
FEATURES	Location/Qualifiers							
source	1. .438							
	/organism="Streptomyces netropsis"							
gene	1. .438	CDS	/db_xref="taxon:55404"					
	/gene="ssgA"							
	1. .>405							
	/gene="ssgA"							
	/note="strain ATCC of Streptomyces netropsis"							
	/codon_start=1							
	/transl_table=11							
	/product="SsgA"		/product="SsgA"					
	/protein_id="CAC07388.1"							
	/db_xref="GI:9995091"							
	/translation="MSFLYSELSFKIPVELRYETRDYPVAVRMTFHLPCDAPVTFWAFG							
	RELLDGINRPSGDGVHIAPDPEGLSDVSRILQVGADRALFRAGAPPLVAFDLDRD							
	KSVPLGQEQTLGDFDSLEAALGKILAEQNAG"							
BASE COUNT	63 a 151 c 131 g 62 t							
ORIGIN	70 a 158 c 142 g 68 t							
Query Match 100.0%; Score 407; DB 6; Length 407;								
Best Local Similarity 100.0%; Pred. No. 2e-38;								
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;								
QY	1	atgagttctctgcgcgagagctctctcaagatcccaatcccaatcgaaactcgatacagag	60					
Db	1	ATGAGCTTCTCTGCTCCGAGGAGCTCTCTTCAAGATCCCACTCGAATCGCATACGAG	60					
QY	61	accgggatactacgcggtgcgagatgacctccacctcccgagagacgcgcctgtgacc	120					
Db	61	ACCGGGATCCCTACGGGTGCGGATGACCTTCCACCTCCCCGGAGACGCGCTGTGACC	120					
QY	121	tgggcttgcgcgagctgtctgcagggatacaaccgccgagcgagcgagcgagcgac	180					
Db	121	TGGGCTTTCGGCGGGAGTGTCTGCTCGACGGGATCAACGCCCGGACGCGCGAC	180					
QY	181	gtccacatgcgcgcgacccgagggcctctcgagcttccatcccgctccaggtg	240					
Db	181	GTCACATGCCCCCGACCCGAGGGCTGTGCGAGCTCTCCATCCGCTCCAGGTG	240					
QY	241	ggcgagacgcgcctcttcctgcagggcgcccgccgctgccttcctccgaccg	300					
Db	241	GGCGGACCGCGCCCTCTTCCTGTCAGGGCGCCCGCGCTGCTCCCTCCGACCGC	300					
QY	301	acggacaagtgcgtgcgcgcgaggaacagactctgggtgacttcgagagacagcctg	360					
Db	301	ACGGACAAGTGCCTGCGCTCGTTCAGGAACAGACTCTGGGTGACTTCGAGGACGCTG	360					
QY	361	gagccgcgctcggaagatcttcgcgagagcagaaacgcgcgctg	407					
Db	361	GAGCCGCGCTCGCAAGATCTTCGCGGAGGACGAGAACCCCGCTG	407					
RESULT	3							
AX007218	AX007218	407 bp	DNA linear PAT 06-SEP-2000					
LOCUS	Sequence 3 from Patent WO0000613.							
DEFINITION								
ACCESSION	AX007218							
VERSION	AX007218.1	GI:9995084						
KEYWORDS	Streptomyces griseus.							
SOURCE	Streptomyces griseus.							
ORGANISM	Streptomyces griseus.							
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces.							
AUTHORS	Kraal, B., Luiten, R.G. and Van Wezel, G.P.							
TITLE	Reducing branching and enhancing fragmentation in culturing filamentous microorganisms							
JOURNAL	Patent: WO 0000613-A 3 06-JAN-2000;							
FEATURES	Location/Qualifiers							
source	1. .407							

/organism="Streptomyces griseus"  
/db\_xref="taxon:1911"  
1. .405  
/gene="ssgA"  
1. >405  
/gene="ssgA"  
/note="strain ATCC of Streptomyces griseus"  
/codon\_start=1  
/transl\_table=11  
/product="SsgA"  
/protein\_id="CAC07385.1"  
/db\_xref="GI:9995083"  
/translation="MSLVSEELSFRIPIVELRYEVDPIAIRMTFHPGADPYTWAFG  
RELLDGLNPSGDGDVHIGTPEPEGLGDVHIRLQVGADRALFRAGTAPLVAFLDRD  
KLVLPGGEHTLIGDFDGNLEADALRAEQNAG"  
BASE COUNT 60 a 139 c 142 g 66 t  
ORIGIN

Query Match 73.3%; Score 298.2; DB 6; Length 407;  
Best Local Similarity 83.3%; Pred. No. 6.5e-26;  
Matches 339; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 1 atgagcttctcttcgcgagagagctctcttcaagatcccgagtcgaactcgatacagag 60  
Db 1 ATGAGCTTCTCTCCGAGGAGCTCTCGTTCCGTATTCCGGTGGAGCTCTGATACGAG 60

Qy 61 accgggacccacgcggtgcgatgacctccacctcccgagagacgcctgtgacc 120  
Db 61 GTCGCGATCCGTATGCCATCCGATGACGTTCCACCTTCCCGGGGATGCCCTGTGACC 120

Qy 121 tggcggttcggcgagagctgctgcacgggatacaaccccgagcgcgacggcgac 180  
Db 121 TGGGGTTTCGGCGGAGCTGCTGCTGACGGGCTCAACAGCCCGAGCGGCGCGAT 180

Qy 181 gtccacatcgccccagccagaccccgagggcctgtcgagctctccatccggctccaggtg 240  
Db 181 GTGCACATCGGCGGAGCGAGCGGCGCTCGGAGATGTCCACATCCGGCTCCAGGTC 240

Qy 241 ggcgcgacgcgcctcttcgtagcagcgcccgccgctggtgcctcttcgacgc 300  
Db 241 GCGCGGACCGTGGCTGTCGGCGGGGACGCGCGCTGGTGCTTCTTCGACCGG 300

Qy 301 acggacaagtcggtcgctcgtagaacagactctgggtgacttcgagagacgcctg 360  
Db 301 ACGGACAGCTCGTCCGCTCGCCAGGACGACAGCTGGTGACTTCGACGGCAACCTG 360

Qy 361 gaggccgcgtcggaagatccttcgcgagagagcagaacgcggctg 407  
Db 361 GAGGACGACTGGGCGCATCTCTGCCGAGGAGCAGACGCGCGGTG 407

RESULT 4  
AX007216  
LOCUS AX007216 438 bp DNA linear PAT 06-SEP-2000  
DEFINITION Sequence 1 from Patent WO0000613.  
ACCESSION AX007216  
VERSION AX007216.1 GI:9995082  
KEYWORDS  
SOURCE Streptomyces griseus.  
ORGANISM Streptomyces griseus  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE  
AUTHORS Kraal,B., Luiten,R.G. and Van Wezel,G.P.  
TITLE 1 (bases 1 to 438)  
JOURNAL Reducing branching and enhancing fragmentation in culturing  
filamentous microorganisms  
PATENT: WO 0000613-A 1 06-JAN-2000;  
UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA  
(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS  
VAN (NL)  
FEATURES Location/Qualifiers  
source 1. .438

/organism="Streptomyces griseus"  
/db\_xref="taxon:1911"  
1. .408  
/gene="ssgA"  
1. >408  
/gene="ssgA"  
/note="strain ATCC of Streptomyces griseus"  
/codon\_start=1  
/transl\_table=11  
/product="SsgA"  
/protein\_id="CAC07384.1"  
/db\_xref="GI:9995083"  
/translation="MRESVQAEVMMSFLVSEELSFRIPIVELRYEVDPIAIRMTFHP  
GDAPVTWAFGRLLDGLNPSGDGDVHIGTPEPEGLGDVHIRLQVGADRALFRAGTA  
PLVAFLDRDRLKVLPGDHTLGDGNLEADALGR"  
BASE COUNT 68 a 145 c 153 g 72 t  
ORIGIN

Query Match 73.3%; Score 298.2; DB 6; Length 438;  
Best Local Similarity 83.3%; Pred. No. 6.3e-26;  
Matches 339; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 1 atgagcttctcttcgcgagagagctctcttcaagatcccgagtcgaactcgatacagag 60  
Db 31 ATGAGCTTCTCTCCGAGGAGCTCTCGTTCCGTATTCCGGTGGAGCTCGGATACGAG 90

Qy 61 accgggacccacgcggtgcgatgacctccacctcccgagagacgcctgtgacc 120  
Db 91 GTCGCGATCCGTATGCCATCCGATGACGTTCCACCTTCCCGGGGATGCCCTGTGACC 150

Qy 121 tggcggttcggcgagagctgctgcacgggatacaaccccgagcgcgacggcgac 180  
Db 151 TGGGGTTTCGGCGGAGCTGCTGCTGACGGGCTCAACAGCCCGAGCGGCGCGAT 210

Qy 181 gtccacatcgccccagccagaccccgagggcctgtcgagctctccatccggctccaggtg 240  
Db 211 GTGCACATCGGCGGAGCGAGCGGCGCTCGGAGATGTCCACATCCGGCTCCAGGTC 270

Qy 241 ggcgcgacgcgcctcttcgtagcagcgcccgccgctggtgcctcttcgacgc 300  
Db 271 GCGCGGACCGTGGCTGCTCGCGGGGAGCGGACCGCTGGTGCTTCTTCGACCGG 330

Qy 301 acggacaagtcggtcgctcgtagaacagactctgggtgacttcgagagacgcctg 360  
Db 331 ACGGACAGCTCGTCCGCTCGCCAGGAGCAGACGCTGGTGACTTCGACGGCAACCTG 390

Qy 361 gaggccgcgtcggaagatccttcgcgagagagcagaacgcggctg 407  
Db 391 GAGGACGACTGGGCGCATCTCTGCCGAGGAGCAGACGCGCGGTG 437

RESULT 5  
D50051  
LOCUS D50051 1513 bp DNA linear BCT 19-MAY-1999  
DEFINITION Streptomyces griseus DNA for ssgA, complete cds.  
ACCESSION D50051  
VERSION D50051.1 GI:1772323  
KEYWORDS ssgA.  
SOURCE Streptomyces griseus (strain:B2682) DNA.  
ORGANISM Streptomyces griseus  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE  
AUTHORS Kawamoto,S. and Ensign,J.C.  
TITLE Cloning and characterization of a gene involved in regulation of  
sporulation and cell division of Streptomyces griseus  
JOURNAL Actinomycetol. 9, 136-151 (1995)  
REFERENCE  
AUTHORS Kawamoto,S. and Ensign,J.C.  
TITLE Isolation of mutants of Streptomyces griseus that sporulate in  
nutrient rich media  
JOURNAL Actinomycetol. 9, 124-135 (1995)



```

/gene="SCQ11.01c"
/complement(<1. .289)
/gene="SCQ11.01c"
/note="SCQ11.01c, hypothetical protein, partial CDS, len:
>95 aa; unknown function, previously sequenced as
TR:Q53870 (EMBL:U37580), S.coelicolor hypothetical protein
ORF5 downstream of phosphotyrosine protein phosphatase
(159 aa). Weak similarity to the N-terminus of TR:Q69888
(EMBL:AL023797). Streptomyces coelicolor hypothetical
protein (172 aa), fasta scores; opt: 158 z-score: 206.4
E(): 0.00036, 46.2% identity in 65 aa overlap. Continues
as SCH24.40c in cosmid H24 (EMBL:AL049826)"
/codon_start=1
/transl_table=11
/label="SCQ11.01c"
/product="hypothetical protein"
/protein_id="CAB46956.1"
/db_xref="GI:5457268"
/transl_table="MTYRVPVVKRTARAVLDGDLHLILIKRTKPGVDPYWTPGGGVPEP
DDTTVDALHREYVEELGAKISDVVPCFVDVTEHIGEDGGATGKVKQHFFVC"
/complement(1. .4279)
/note="sequence corresponding to EMBL:U37580 from 1 to
4279"
1. .98
/note="overlap with Streptomyces coelicolor cosmid H24
(EMBL:AL049826) from 41528 to 41625"
/complement(360. .1265)
/gene="SCQ11.02c"
/note="abaB"
/complement(360. .1265)
/gene="SCQ11.02c"
/note="SCQ11.02c, abaB, probable lysR-family
transcriptional regulator, len: 301 aa; previously
sequenced as TR:Q53869 (EMBL:U37580) S.coelicolor
lysR-like protein (301 aa). Highly similar to
SW:ARAB_STRAT (EMBL:X91393), araB, Streptomyces
antibioticus transcriptional regulatory protein (301 aa),
fasta scores; opt: 1581 z-score: 1813.2 E(): 0, 80.3%
identity in 300 aa overlap. Similar to many other
transcriptional regulators. Contains probable
helix-turn-helix motif at aa 17-38 (Score 1659, +4.84 SD).
Contains Pfam match to PF00126 HTH_1, Bacterial regulatory
helix-turn-helix protein, lysR family and PS00044
Bacterial regulatory proteins, lysR family signature"
/codon_start=1
/transl_table=11
/label="abaB"
/product="putative lysR-family transcriptional regulator"
/protein_id="CAB46957.1"
/db_xref="GI:5457269"
/transl_table="MDLALLRTFVTVHRAGSFTRAAALLGLSOPAVTSQIRTLRLQGLG
RPLFLRQAGVPTTIGDELAKHAAPHLDALVETAEGLDEDDSTLRLHLAGPEFTA
ERALPALGELTGEDGQAFALRASFGNAETLEGLAAGHDLAIGTTTPRGALHTATPL
CDEVHLVATPHWAERAGVEDVTDASALKHVPVEVHESLPFVGRYWASVFADAPA
SPATVAVLRAVLACAVAGALVLAFLRYLCAEALERGDVVVALHDPVPVPLRTVFLVV
RTGTLPAPHIARAHEWLLRAAADWN"
/complement(831. .1256)
/gene="SCQ11.02c"
/note="Match to PF00126 HTH_1, Bacterial regulatory
helix-turn-helix protein, lysR family Score 145.96"
/complement(1125. .1217)
/gene="PS00044 Bacterial regulatory proteins, lysR family
signature"
/complement(1518. .2696)
/gene="SCQ11.03c"
/note="cysA"
/complement(1518. .2696)
/gene="SCQ11.03c"
/note="SCQ11.03c, cysA, probable cystathionine/methionine
gamma-synthase/lyase, len: 392 aa; previously sequenced as
SW:CYS_A_STRCO (EMBL:U37580), cysA, S.coelicolor putative
cystathionine gamma-lyase (392 aa). Similar to many e.g.

```

---

```

SW:MEGL_PSEPU (EMBL:D88554), mdeA, Pseudomonas putida
methionine gamma-lyase (398 aa), fasta scores; opt: 608
z-score: 653.6 E(): 4.5e-29, 36.9% identity in 404 aa
overlap. Highly similar to TR:Q53668 (EMBL:X91393)
Streptomyces antibioticus hypothetical protein found
upstream of the abaB gene (232 aa) (87.1% identity in 232
aa overlap). Contains Pfam match to PF01053
Cys_Met_Meta_PP, Cys/Met metabolism PLP-dependent enzyme
and PS00868 Cys/Met metabolism enzymes pyridoxal-phosphate
attachment site"
/codon_start=1
/transl_table=11
/label="cysA"
/product="putative cystathionine/methionine
gamma-synthase/lyase"
/protein_id="CAB46958.1"
/db_xref="GI:5457270"
/transl_table="MSDSATTSAGTGGERSASAPGDGTRAVRAGLPVPVKEPTLPG
PVFAAHFHPGDPTGPTTYGRDENPTWTRLESAIGEAPGEAGVETLVFASGMAAIS
SVLFSQLRAGDTAVLPDDGYOALPLVRAQLEAYGIEVTAFTPGDAQDLVDGAKLLW
IETSPCLDVCVDRRLVEAAHAGGALVAVDNTLATPLGQRPLEIGADFSVSGTKOL
TGHGDVLLGYVAGRDAGMAAVRWKTVGAIPGPMENLAHRSATLQLKVRDQDST
ALKVAELKTRPEITGLRYPGLPDDPSHKVASQQLRYGCVVSTFLPSRALRDLDA
LRLVEGATSGGVRSTAEERRGRWGDDAVPEGFIRLSVGAEDPDDLVALDLDETTE
"
/complement(1533. .2624)
/gene="SCQ11.03c"
/note="Match to PF01053 Cys_Met_Meta_PP, Cys/Met
metabolism PLP-dependent enzyme Score 268.52"
/complement(2031. .2075)
/gene="SCQ11.03c"
/note="PS00868 Cys/Met metabolism enzymes
pyridoxal-phosphate attachment site"
/complement(2693. .3187)
/gene="SCQ11.04c"
/note="ptpA"
/complement(2693. .3187)
/gene="SCQ11.04c"
/note="SCQ11.04c, ptpA, low molecular weight
protein-tyrosine-phosphatase, len: 164 aa; previously
sequenced as SW:PTPA_STRCO (EMBL:U37580), ptpA,
S.coelicolor low molecular weight
protein-tyrosine-phosphatase (164 aa). Contains Pfam match
to PF01451 LMWPC, Low molecular weight phosphotyrosine
protein phosphatase"
/codon_start=1
/transl_table=11
/label="ptpA"
/product="low molecular weight
protein-tyrosine-phosphatase"
/protein_id="CAB46959.1"
/db_xref="GI:5457271"
/transl_table="MTYRVCFVTCNCRSPMAEAVFARVEDAGLHGLVADSAGTG
GWHBEGADPRTEALADHGVLDAARFOQSWFRLDLVLVADLALRLRLAPT
ERDAAKVRLRLRSYDPAVAGGLDVPDPYVGGRGDFECLLENVEASTGLLAAREQVE
GRAA"

```

---

```

Query Match      62.3%; Score 253.4; DB 1; Length 15441;
Best Local Similarity 76.4%; Pred. No. 3.3e-21;
Matches 311; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 1 atgagcttcctcgctccgagagctctctccatccagatcccaactcgactcgatacgaag 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5475 ATGAGCTTTCGCTCCGAGGAGGCTCTCTTCCCATCCCGGAGACGCGCGGTGACC 5534

Qy 61 acccgggatccctacgcggtgcgagatgaccttccacctcccccgagagcgcgtgacc 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5535 ACCCGGATCCCTATGCCGTACGCCCTGACCTTTCATCTGCCCGGAGACGCGCGGTGACC 5594

Qy 121 tggcggttcggccggagctgctgcacgggatacaaccgcccgagcgagcgacggc 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5595 TGGGCTTCGCGCGGAGCTGCTGTGCACGCGGTGGCGCGCGCTGCGCGCACGCGAC 5654

```



```
Db 121 TGGCGTTCGGCGGGAGTTCTCATCGACGAGGTCCGCGCGCTGCGGGAGCGGGAC 180
Qy 181 gtccacatcgcccgaccgagccgagcctgtcgagcgtctccatccggtccaggtg 240
Db 181 GTCCACATCGCGCGCGCGGAGAGAGTTCGGCGAGGTCCTGATCCGCTCGAGGTG 240
Qy 241 ggcgagaccgcgccctcttcctgcagggcgcccgccgctggctccttcgcagcgc 300
Db 241 GGGAGGACGACGAGCGATGTTCCGGGTGCGACGCGCGCTGGTGGCTTCTTGGACGC 300
Qy 301 acgagacaagtgcgctcgcgctcgcgctcgcgctcgcgctcgcgctcgcgctcgc 360
Db 301 ACGGACACAAGATCGTCCCGCTGCGGCGAGGAGGTTCCCTCGCGGACTTTCGACGCTGCTC 360
Qy 361 gaggcgcgctcgcaagatcctcgccgagagcagacgagcgcgctg 407
Db 361 GACGAGGCGCTGGACCGCATCTGCGCGAGGAGGAGACGCGCGGTG 407

RESULT 9
AX007220
LOCUS AX007220 407 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 5 from Patent WO0000613.
ACCESSION AX007220
VERSION AX007220.1 GI:9995086
KEYWORDS
SOURCE Streptomyces albus G.
ORGANISM Streptomyces albus G.
REFERENCE
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 407)
AUTHORS Kraal,B., Luiten,R.G. and Van Wezel,G.P.
TITLE Reducing branching and enhancing fragmentation in culturing
filamentous microorganisms
JOURNAL Patent: WO 0000613-A 5 06-JAN-2000;
UNIV LEIDEN (NL); KRAAL BAREND (NL); LUITEN RUDOLF GIJSBERTUS MARIA
(NL); NL ORGANISATIE VOOR WETENSCHAP (NL); WEZEL GILLES PHILIPPUS
VAN (NL)
FEATURES
source Location/Qualifiers
1..407
/organism="Streptomyces albus G"
/db_xref="taxon:1962"
1..405
/gene="ssgA"
1..>405
/gene="ssgA"
/strain ATCC of Streptomyces albus G"
/codon_start=1
/transl_table=11
/product="SsgA"
/protein_id="CAC07386.1"
/db_xref="GI:9995087"
/translation="MSFLVSEELAFRIPVELRYETDPYAVRLTFHLPGBAPVTWFG
RELLVEGLDAGDVRVCPVGQTATREVHITLQVGEALFRVGRKAPLLAFLDRTD
QGLSLGERAHADFDHLDALNRSLAEQASG"
BASE COUNT 57 a 150 c 139 g 61 t
ORIGIN

Query Match 55.2%; Score 224.6; DB 6; Length 407;
Best Local Similarity 72.0%; Pred. No. 1.9e-17;
Matches 293; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 1 atgagcttctcgttcgcgagagctctcctcaagatccccagtcgaactcgatacag 60
Db 1 ATGAGCTTCTCTCGTTCGCGAGGAGTCCGCTTCGCATCCCGGTGGAGCTCGGTACGAG 60
Qy 61 accggggtccctacgcggtcgagatgacettccacctccccggcgagagcgccgtgacc 120
Db 61 ACCGTCGATCCGTACGCGGTGCGGCTGACGTTCACACCTCCCGGAGACGCGCGGTACC 120
Qy 121 tgggcgttcgagcgagctgctgctgacgaggatcaaccgcccgcgagcgcgagcgagac 180
```

```
Db 121 TGGGTCTTCGGCGCGTGAACCTGCTGTCGAGGAGTCTCTGGACGCGCGGGCGACGCGAC 180
Qy 181 gtccacatcgcccgaccgagcccgagcctgtcgagcgtctccatccggtccaggtg 240
Db 181 GTCCGGGTCTGCCCGGTGGGAGAGCGCCACACGAGGAGGTGCACATCACCTTCAGGTG 240
Qy 241 ggcgagaccgcgccctcttcctgcagggcgcccgccgctggctccttcgcagcgc 300
Db 241 GGCCTCGAGCAGGCGCTCTTCCGGGTGCGCAAGGCGCGCTGCTCGCTTCTCTCGACGC 300
Qy 301 acgagacaagtgcgctcgcgctcgcgctcgcgctcgcgctcgcgctcgcgctcgc 360
Db 301 ACCGACCAAGGCTTGTGCTGTCGACGAGCGGGGACACGCGGACTTCGACAGCCACCTC 360
Qy 361 gaggcgcgctcgcaagatcctcgccgagagcagacgagcgcgctg 407
Db 361 GACGAGGCTCTGAACCGCAGGCTCGCCGAGGAGGAGCAGCGCGGTG 407

RESULT 10
AF195771
LOCUS AF195771 566 bp DNA linear BCT 01-NOV-2000
DEFINITION Streptomyces albus SsgA (SsgA) gene, complete cds.
ACCESSION AF195771
VERSION AF195771.1 GI:11066160
KEYWORDS
SOURCE Streptomyces albus.
ORGANISM Streptomyces albus.
REFERENCE
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 566)
AUTHORS van Wezel,G.P., Rousseau,C. and Kraal,B.
TITLE Cloning and sequencing of the Streptomyces albus ssgA gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 566)
AUTHORS van Wezel,G.P.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Biochemistry, University of Leiden, PO Box
9502, Leiden 2300 RA, Netherlands
FEATURES
source Location/Qualifiers
1..566
/organism="Streptomyces albus"
/db_xref="taxon:1888"
16..433
/gene="ssgA"
16..20
/gene="ssgA"
26..433
/gene="ssgA"
/strain similar to Streptomyces griseus SsgA; possible cell
division protein"
/codon_start=1
/transl_table=11
/product="SsgA"
/protein_id="AAG28482.1"
/db_xref="GI:11066161"
/translation="MSFLVSEELAFRIPVELRYETDPYAVRLTFHLPGBAPVTWFG
RELLVEGLDAGDVRVCPVGQTATREVHITLQVGEALFRVGRKAPLLAFLDRTD
QGLSLGERAHADFDHLDALNRSLAEQASG"
BASE COUNT 75 a 219 c 192 g 80 t
ORIGIN

Query Match 55.2%; Score 224.6; DB 1; Length 566;
Best Local Similarity 72.0%; Pred. No. 1.7e-17;
Matches 293; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 1 atgagcttctcgttcgcgagagctctcctcaagatccccagtcgaactcgatacag 60
Db 26 ATGAGCTTCTCTCGTTCGCGAGGAGTCCGCTTCGCATCCCGGTGGAGCTCGGTACGAG 85
Qy 61 accggggtccctacgcggtcgagatgacettccacctccccggcgagagcgcgctgtgacc 120
```



Db	86	ACCGTCGATCCGTCGCGGTGAGCTTCACCTCCCGGAGACGCCCGGTCACC	145
Qy	121	tgggcttcggccggagctgctgctgcagcgagatacaaccgcccagcgacgacgacgac	180
Db	146	TGGGCTCTCGGGCGTAAGTCTGCTGAGGAGTCTTGAGACCCCGCGGACGCGAC	205
Qy	181	gtcacatcgcccgaccgaccccgagggcctgtgctggagctgtccatcggtctccaggtg	240
Db	206	GTCCGGTCTGCCGCTGGCGGACAGCGCCACCAGGAGGTGCACATCACCTCCAGGTC	265
Qy	241	ggcggaacggccctcttcgctgagcgcccccgccgctggtgcctctctccagcgc	300
Db	266	GGCTCCGAGCAGCGCTCTTCCGCGCTCGCAAGCGCCGCTGCTCCCTTCCTCGACCG	325
Qy	301	acgacaagtgcgctgcgctcgtcagcacagactcgtggtgacttcgagacagcctg	360
Db	326	ACCGCAGGGCTTGCTGCGTCGACGCGGCGGACACGCCGACATTCACACCGCACCTC	385
Qy	361	gagcgcgctcggaagatctctcgccgagggagcagaacgcggctg	407
Db	386	GACGACGCTCTGAACCGCAGCCTCGCGGAGGACGAGCGCGGCTG	432
RESULT	11		
LOCUS	SCL2	38640 bp	DNA linear BCT 29-JAN-2000
DEFINITION	Streptomyces coelicolor cosmid L2.		
ACCESSION	AL137778		
VERSION	AL137778.1	GI:6822206	
KEYWORDS	acyltransferase; acyltransferase; adenine phosphoribosyltransferase; DNA helicase; DNA polymerase III; elongation factor; endonuclease; GTP pyrophosphokinase; membrane protein; membrane transferase; protein-export membrane protein; regulator; secreted protein; sugar transferase; threonine-tRNA synthetase; transport system inner membrane protein. Streptomyces coelicolor A3(2).		
SOURCE	Streptomyces coelicolor A3(2)		
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
REFERENCE	1 (bases 1 to 38640)		
AUTHORS	Redenbach, M., Kieser, H.M., Denapalte, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D.A.		
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome		
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)		
MEDLINE	97000351		
REFERENCE	2 (bases 1 to 38640)		
AUTHORS	Seeger, K.J. and Harris, D.		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 38640)		
AUTHORS	Cerdeno, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.		
TITLE	Submitted (28-JAN-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge		
JOURNAL	CB10 ISA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK		
COMMENT	Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: <a href="http://www.sanger.ac.uk/Projects/S_coelicolor/">http://www.sanger.ac.uk/Projects/S_coelicolor/</a> ) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given		

where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid L2.

FEATURES	Location/Qualifiers	source
gene	1..38640 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid L2"	
CDS	1..989 /gene="SCL2.01" /partial	
gene	1..989 /note="SCL2.01" /note="SCL2.01, hypothetical protein, len: >328 aa; similar to TR:O33336 (EMBL:298209) Mycobacterium tuberculosis hypothetical 49.8 kD protein MTCV174.11, len: 450 aa; fasta scores: opt: 843 z-score: 874.7 E(0); 0; 45.9% identity in 296 aa overlap. Contains possible coiled-coils region"	
CDS	1010..1150 /codon_start=3 /transl_table=11 /product="hypothetical protein SCL2.01" /protein_id="CAB70913.1" /db_xref="GI:6822207"	
gene	1010..1150 /gene="SCL2.02" /note="SCL2.02, unknown, len: 46 aa"	
CDS	1010..1150 /codon_start=1 /transl_table=11 /product="hypothetical protein SCL2.02" /protein_id="CAB70914.1" /db_xref="GI:6822208"	
misc_feature	complement(1218..3934) /note="previously sequenced region SW:SCAPTRELA EMBL:X87267 S.coelicolor apt & rela genes"	
gene	complement(1229..3772) /gene="relA"	
CDS	complement(1229..3772) /gene="relA" /note="SCL2.03c, relA, GTP pyrophosphokinase, len: 847 aa; identical to previously sequenced SW:RELA_STRCO (EMBL:X87267) Streptomyces coelicolor GTP pyrophosphokinase (EC 2.7.6.5) RelA, 847 aa and highly similar to SW:RELA_STRAT (EMBL:AF072829) Streptomyces antibioticus RelA protein, 841 aa; fasta scores: opt: 4969 z-score: 4906.2 E(0); 90.3% identity in 848 aa overlap and to SW:RELA_CORGL (EMBL:AF038651) Corynebacterium glutamicum GTP pyrophosphokinase (EC 2.7.6.5) RelA, 760	



```
aa; fasta scores: opt: 1613 z-score: 1594.3 E(): 0; 58.7%
identity in 774 aa overlap. Contains Pfam match to entry
PF01842 ACT, ACT domain and match to Prosite entry PS00017
ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/transl_table=11
/product="GTP pyrophosphokinase"
/protein_id="CAB70915.1"
/db_xref="GI:6822209"
/translacion="MPDEAQLTAAPKESASAKPAPSAPOAKNDTHGPIHAPAA
PVDPAPQPPKPLPAERQNPAPVAPAGCAPQARSNNRVARLARGVORANPIN
PVLPLLRIVAGNDPKLLETSTLROIERAYQVAERHGRQKRGSDPTITHPLAVTTL
AELGMDPATLWGLDHTDETEYLEDLRDFGVVTLVLDGVTKLDKFKFGAAQA
ETVRKMYAKADPRVLVVKMLADLRHNNRMRYLREKQKARETEIYAPLAHLRG
MNTLKWELEDAPAILVPKMDVETRLVAERAPKRDYLAVTVDEVDLRAARIKAT
VTGPKHYVSYVQKMIWGRDFAEIVDLGVLRLVDTVRDCYALGTVHARWNPVGR
FKDIAMPKNNYQSLHTTVYFGPKFVELQIRTFDMHRAEYGAIAHWKYKQAVAG
ASKVTDAPKSGKSDHNDMAWLQDQKEDTDPGEFLESRLDSRNVFVF
TPKGDVIALPAGATPVDFAYAVHTVEVGHRTIGARVNGRLVPLESTLNDGLVEYFTSK
AAGAPSRDMLGFVKSPPARKIRAFWFSKERRDEAIEQKDAIVRAMRKQNLPTQRI
TGDLSLVLAMHRYSDISALYAAIGEGHVSAPNTVOKLVOALGGEAEATEIDSVPP
SRGGRKRRANADPGVVKGVEDVWVKLARCCTVPVGDPIIGFTVTRGSGVHRSDCV
NVDSLSREPERILEVNAQTSSVFLVAIQEALDRSLSDVTRVLSQDVHNLISCA
VQTSRDRVATSRFTFEMGDPKHLGHVILKAVRGVEGVYDVYRVTSARRPS"
complement(1256..1474)
/gene="rela"
/notes="Pfam match to entry PF01842 ACT, ACT domain, score
63.10, E-value 6e-15"
complement(2378..2401)
/gene="rela"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(3273..7460)
/notes="previously sequenced region SW:SCSECAPT EMBL:X85969
S.coelicolor secB, secF & apt genes"
complement(3778..3781)
complement(3904..4568)
/notes="previously sequenced region SW:SCAPTRELA
EMBL:X87267 S.coelicolor apt & relA genes"
complement(3956..4453)
/gene="apt"
complement(3956..4453)
/gene="apt"
/notes="SCL2.04c, apt, adenine phosphoribosyltransferase,
len: 182 aa; identical to previously sequenced
SW:APT_STRCO (EMBL:X87267) Streptomyces coelicolor adenine
phosphoribosyltransferase (EC 2.4.2.7) Apt, 182 aa and
highly similar to SW:APT_ECOLI (EMBL:M14040) Escherichia
coli adenine phosphoribosyltransferase (EC 2.4.2.7) Apt,
183 aa; fasta scores: opt: 600 z-score: 682.8 E():
1.2e-30; 54.8% identity in 168 aa overlap. Contains Pfam
match to entry PF00156 Pribosyltran, Phosphoribosyl
transferase domain and match to Prosite entry PS00103
Purine/pyrimidine phosphoribosyl transferases signature"
/codon_start=1
/transl_table=11
/product="adenine phosphoribosyltransferase"
/protein_id="CAB70916.1"
/db_xref="GI:6822210"
/translacion="MADYPEGVVFKDITPLLDPGAPAAITDALAEAGRTGATKV
GLEARGFILGAPVALRAGLGPVPRKAGKLPATLSQAYDLEYSAEIVHAEDLTAG
DRVLVDDVLATGCTAEASLELIRRAEAGVAGLAVLMELGLFGLGRARLEPALAGAPLE
ALLTV"
complement(3980..4429)
/gene="apt"
/notes="Pfam match to entry PF00156 Pribosyltran,
Phosphoribosyl transferase domain, score 150.30, E-value
3.5e-41"
complement(4103..4141)
/gene="apt"
/notes="PS00103 Purine/pyrimidine phosphoribosyl
transferases signature"
complement(4468..4503)
complement(4501..5622)
```

```
/gene="secF"
complement(4501..5622)
/gene="secF"
/notes="SCL2.05c, secF, protein-export membrane protein,
len: 373 aa; identical to previously sequenced
SW:SECF_STRCO (EMBL:X85969) Streptomyces coelicolor
protein-export membrane protein SecF, 373 aa and similar
Query Match 28.2%; Score 114.8; DB 1; Length 38640;
Best Local Similarity 55.9%; Pred. No. 2.2e-05;
Matches 218; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
QY 2 taagcttcctctccgaggagctctctcaagatccacagtcgcaactcgcatcacgaga 61
Db 35030 TCGCGCTCGTTGTGTCGAGCGAGTCTCCCTGCTGTCCTCCCGCAGGCGCTCGGTACGACA 35089
QY 62 cccgggataccctacggtgtcggtatgacattccacattcccccggagacgctgtgacct 121
Db 35090 CGCGCGACCCCTACGCGGTGTCACGGCCACCTTCACACCGGAGCGGAGACCGTGCAGT 35149
QY 122 gggccttcggtcgagctgctcgacggatcaaccccgagcgcgcgacgagcgagc 181
Db 35150 GGGTCTTCGCCCGCACCTGCTCGCGAAGGTCTCCACCGTCCACCGGACCGCGGACG 35209
QY 182 tcacatgccccgacccgagggcctgtcgagctctccatcccggtccaggttg 241
Db 35210 TCCGTGTCTGGCGGTCCCGCAGTCAGGCGCAGGGCGTGTGTGCATCGCTCAGCTCCC 35269
QY 242 gcgcggacgcgcctcttcctgcaggcgcccgccgctgtgctgcttccttcgacgcga 301
Db 35270 CGGAGGGCGAGGACTGCTCGAGGCGCGCGCGCGGCGCTGAGTCTCTTCTTGAAGCGCA 35329
QY 302 cggacaagtgcgtcgctgcaggaacagacttgggtgacttcgagggacgagcctgg 361
Db 35330 CAGACGCCCGCTGCCCGCCGACGGAACCGGCACCTTCGATCTCGATCAGGAGCTCT 35389
QY 362 aggcgcgcctgcgaagatcctcgccgag 391
Db 35390 CGCATCTCTGGCGGNAAGCTAGGCGGGG 35419
RESULT 12
LOCUS SCE19A 35284 bp DNA linear BCT 16-JUL-1999
DEFINITION Streptomyces coelicolor cosmid E19A.
ACCESSION AL096852
VERSION AL096852.1 GI:5531349
KEYWORDS 4-hydroxyphenylpyruvate dioxygenase; ABC transporter; amino acid
permease; asc-family; cysteine synthase; nicotinamidase;
osmoregulation; osmotolerance; oxidoreductase; permease;
phosphoenolpyruvate-dependent; protease; regulator; ribonuclease
PH; rph; sugar phosphotransferase system; transcriptional
regulator; transposase; tRNA Leu; zinc metalloproteinase.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 35284)
AUTHORS Redenbach,M., Kieser,H.M., Denapate,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
JOURNAL 97000351
MEDLINE 2 (bases 1 to 35284)
REFERENCE Seeger,K. and Harris,D.
AUTHORS Unpublished
JOURNAL 3 (bases 1 to 35284)
AUTHORS James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
```

David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.

(URL: [http://www.sanger.ac.uk/projects/S\\_coelicolor/](http://www.sanger.ac.uk/projects/S_coelicolor/))

CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.

The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid E19A lies between and overlaps cosmids E6 and E56 on the AseI-E genomic restriction fragment.

Location/Qualifiers

1..35284

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clones="cosmid E19A"

1..97

/gene="SCE19A.01"

<1..97

/note="SCE19A.01"

/note="SCE19A.01, hypothetical protein, partial CDS, len: >31 aa; unknown function, similar to members of the alkyl hydroperoxide reductase C/thiol-specific antioxidant family e.g. Tr:053226 (EMBL:AL021185), bcp, Mycobacterium tuberculosis bacterioferritin comigratory protein (157 aa), fasta scores; opt: 124 z-score: 238.1 E(): 6.2e-06, 63.3% identity in 30 aa overlap"

/codon\_start=2

/transl\_table=11

/label="SCE19A.01

/product="hypothetical protein"

/protein\_id="CAB50982.1"

/db\_xref="GI:5531350"

/translation="IVVDEQKVERALYNNRANGHVAKIKDLGI"

140..226

/note="trna Leu anticodon TAG, Cove score 69.78"

/product="trna-Leu"

complement(243..845)

/gene="SCE19A.02c"

complement(243..845)

/note="SCE19A.02c"

/note="SCE19A.02c, conserved hypothetical protein, len: 200 aa; unknown function, similar to many e.g. SW:Y03Q\_MYCTU (EMBL:Z73902) Mycobacterium tuberculosis hypothetical protein (204 aa), fasta scores: opt: 633 z-score: 743.6 E(): 0.52.0% identity in 200 aa overlap. Similar to SW:YGVG\_ECOLI (EMBL:U28377), yggv, Escherichia coli hypothetical protein (197 aa) (49.5% identity in 198

aa overlap)"

/codon\_start=1

/transl\_table=11

/label="SCE19A.02c

/product="hypothetical protein"

/protein\_id="CAB50983.1"

/db\_xref="GI:5531351"

/translation="MTRLILATRNAGKITELRALADAGLPHDLVGDADYPHIPDVKE TGVTFANALLKAHALEATGLPAVADDSGLCVDLNACPGIF SARWAGRHGDDQANL DLLAQIADTADDEHRAHFACAAALALPDLPTERVVEGOLKGTLRHAPAGTGGFGYDPI LQPEGETRTCAELTAEEKNALSHRKAFLALPVVYRELIG"

complement(892..1296)

/gene="SCE19A.03c"

complement(892..1296)

/note="SCE19A.03c, possible secreted protein, len: 134 aa; unknown function, probable CDS suggested by GC frameplot, positional base preference and amino acid composition. Contains probable N-terminal signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site, so may be a lipoprotein"

/codon\_start=1

/transl\_table=11

/label="SCE19A.03c

/product="putative secreted protein"

/protein\_id="CAB50984.1"

/db\_xref="GI:5531352"

/translation="MAASRRRLRRRTVATVATIALTAGLTTCDAVKALDCVRT ADAIDSVTLEQAVENADDPQWEESLNSIDKNLRIGDTNTDYNKAVDDILGKAV DNVRTSVENGDETPDLSPVTDAAAGELTKVCTP"

complement(1201..1233)

/gene="SCE19A.03c"

/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"

complement(1454..2191)

/gene="SCE19A.04c"

/note="rph"

complement(1454..2191)

/gene="SCE19A.04c"

/note="SCE19A.04c, rph, probable ribonuclease PH, len: 245 aa; highly similar to many e.g. SW:RNPH\_ECOLI (EMBL:X00781), rph, Escherichia coli ribonuclease PH (238 aa), fasta scores; opt: 916 z-score: 1116.8 E(): 0.60.8% identity in 240 aa overlap. Contains pfam match to entry PF01138 RNase\_PH, 3' exonuclease family and PS01277 Ribonuclease PH signature"

/codon\_start=1

/transl\_table=11

/label="rph

/product="putative ribonuclease PH"

/protein\_id="CAB50985.1"

/db\_xref="GI:5531353"

/translation="MSRIDGRTPOQLRPVTIERGWSKHAEGSVLVSFGDKVLCNASV TEGVPRWRKSGEGWTAETAYMLPRATNTRGDSVKGRIGRTHETSLRIGSLRAV IDYKALGENTVLDGCVLOADGRTTAATGAYVALADAVAGQKRLIKANRKLPTG TVSAVSIGIVDGTPLLDLRYEEDVRADTDNNVCTGDSREVEVQGTAEAEFFARDELN TLIDLATAGCTELAELOKALDATALER"

complement(1511..2185)

/note="SCE19A.04c"

/note="Pfam match to entry PF01138 RNase\_PH, 3' exonuclease family, score 266.70, E-value 3.1e-76"

complement(1805..1843)

/gene="SCE19A.04c"

/note="PS01277 Ribonuclease PH signature"

complement(2268..2501)

/gene="SCE19A.05c"

complement(2268..2501)

/note="SCE19A.05c"

/note="SCE19A.05c, hypothetical protein, len: 77 aa; unknown function, similar to parts of many sugar permeases e.g. TR:P96159 (EMBL:U05013), malX, Vibrio furnissii PTS (phosphoenolpyruvate-dependent sugar phosphotransferase

## COMMENT

## FEATURES

## gene

## CDS

## tRNA

## gene

## CDS

system) permease for glucose (523 aa), fasta scores; opt: 217 z-score: 287.9 E(): 1.1e-08, 48.1% identity in 77 aa overlap. Contains Pfam match to entry PF00367 PTS\_EIIB, phosphotransferase system, EIIB"

/codon\_start=1  
/transl\_table=11  
/label=SCE19A.05c  
/product="hypothetical protein"  
/protein\_id="CAB50986.1"  
/db\_xref="GI:5531354"  
/translation="MASKAEKIVAGLGIDNIDEIGCTRLRTEVNDPALVNEAALK AAGAHGVKMGTAIQVIGTDADPIAAEIDMM"  
complement(2385..2489)  
/gene="SCE19A.05c"  
/note="Pfam match to entry PF00367 PTS\_EIIB, phosphotransferase system, EIIB, score 43.00, E-value 8.1e-11"  
2660..3955  
/gene="SCE19A.06"  
2660..3955  
/gene="SCE19A.06"  
/note="SCE19A.06, possible PTS transmembrane component, len: 431 aa; similar to many PTS (phosphoenolpyruvate-dependent sugar phosphotransferase

misc\_feature

Query Match 26.4%; Score 108.2; DB 1; Length 35284;  
Best Local Similarity 55.4%; Pred. No. 0.00013;  
Matches 209; Conservative 0; Mismatches 168; Indels 0; Gaps 0;  
QY 2 tgaggtctcttcgagagctctctcaagatccagtcgaactcgatcgatcgaga 61  
Db 20142 TCAGGCTCGTCTGTCGCCGAGAGCGGCATTCGCGTGCCGCCCGCTCGGCTACCA 20201  
QY 62 cccggagatccatcggtgagatgacattccactccccggagagcgctgtgacct 121  
Db 20202 CGGACGATCCCTACCGCTGCACATCCATCCGATCCGCGCCACCGGTCGCACT 20261  
QY 122 gggcgttcggcgagctgctgctcgacgggatcaacccgcccggcgagcgagcgagc 181  
Db 20262 GACGTTGCGCGCGACCTTCTGTTGAGGGCGTCTTCGCGCGCTCGCGGACGGGAGC 20321  
QY 182 tccatcgccccaccgacccggcgctgctggaogtctccatccgggtccaggtgg 241  
Db 20322 TCGCGGTGTGGCGTCCGAAGACGAGGCGCGCAGCGTCTACTCGTCCGCTGAGCAGCC 20381  
QY 242 gcgagacggcgccctcttcgagcgcccgccgctgctgctcccttcgacccgca 301  
Db 20382 CGGACGGCGACGCCCTCTGTGAGGCGCCACGCCCCAGGTGTGCGCCTGGGTGGAGCGGA 20441  
QY 302 cggacaagtcggtcgctcggtgaggaacagactctggtgacttcgagggacagcctgg 361  
Db 20442 CCTCGGGCGGTGCGCGGGGACCGGCGCGCAGCTCGGGATCGACGAGCGGTGG 20501  
QY 362 aggcgcgcgtcggaag 378  
Db 20502 CCGAGCTGCTGCCAGG 20518

RESULT 13  
STMLACBG  
LOCUS 1410 bp DNA linear BCT 26-APR-1993  
DEFINITION S.albus beta-lactamase gene, complete cds.  
ACCESSION M28303  
VERSION M28303.1 GI:153338  
KEYWORDS beta-lactamase.  
SOURCE Streptomyces albus (strain G) DNA, clone pDML6.  
ORGANISM Streptomyces albus  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 1410)  
REFERENCE Dehottay,P., Dusart,J., De Meester,F., Joris,B., Van Beeumen,J.,  
AUTHORS Epicum,T., Frere,J.-M. and Ghuyssen,J.-M.  
TITLE Nucleotide sequence of the gene encoding the Streptomyces albus G

beta-lactamase precursor  
Eur. J. Biochem. 166, 345-350 (1987)  
87275916 Location/Qualifiers  
1..1410  
/organism="Streptomyces albus"  
/strain="G"  
/db\_xref="taxon:1888"  
243..1242  
/gene="beta-lactamase"  
243..1187  
/gene="beta-lactamase"  
/codon\_start=1  
/transl\_table=11  
/product="beta-lactamase"  
/protein\_id="AAA26775.1"  
/db\_xref="GI:153339"  
/translation="MHPSTRSPRRTLTATAGAALAAATLVPGTAHASSGGRGHSG SVSDAERRLAGRLASGARLVGYAYDTGSGRTVAYRADELFPMCSVFVTLSSAALVLRD LDRNGEFLSRRLTYTDDVEQADGAGPETGKPNLANAQLTVEELCEVSTIASDNCAA NMLRELGGPRAVTRFVSLGDRVTRLDWEPELNSAEPGRVTDTSPPRAITRTGRL VLGDALNPRDRLLTSWLLANTTSGDRFRAGLPDDWTGLGDKTGAGRYGTNNDAVGTWP PGRAPIVLTVLFAKTEQDAARDGLVADAARVLAETLG"

repeat\_unit 196 a 540 c 496 g 178 t  
ORIGIN

Query Match 23.4%; Score 95.4; DB 1; Length 1410;  
Best Local Similarity 81.0%; Pred. No. 0.0095;  
Matches 111; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 22 gagctctctcaagatccagtcgacgacgacgacgacgacgacgacgacgacgacg 81  
Db 1 GAGCTCGCTTCGCGATCCGCTGGAGCTCGGTAGACACCGCTCGATCGTACGCGGTG 60  
QY 82 cggatgaccttcacactccccggagagcgcgctgtgacctggcgctcgccgggagctg 141  
Db 61 CGGCTGACGCTTCACCTCCCGGAGAGCGCCCGCTCACCTGGGTCTTCGGGCGTGA 120  
QY 142 ctgctcgacggagatcaa 158  
Db 121 CTGGTCGAGGGAGTCAA 137

RESULT 14  
SC8A11  
LOCUS 32704 bp DNA linear BCT 04-AUG-2000  
DEFINITION Streptomyces coelicolor cosmid 8A11.  
ACCESSION AL391041  
VERSION AL391041.1 GI:9716211  
KEYWORDS amino acid transport integral membran protein; branched amino acid binding secreted protein; branched amino acid transport system permease; ATP-binding protein; branched amino acid transport integral membrane protein; integral membrane protein; ion transport integral membrane protein; polysaccharide biosynthesis protein; regulator; regulatory protein; secreted amidase; secreted peptidase; secreted protein; sigma factor; transcriptional regulator; transcriptional regulatory protein.  
SOURCE Streptomyces coelicolor A3(2).  
ORGANISM Streptomyces coelicolor A3(2)  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 32704)  
REFERENCE Redenbach,M., Kieser,H.M., Denapate,D., Eichner,A., Cullum,J.,  
AUTHORS Kinashi,H. and Hopwood,D.A.  
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351  
2 (bases 1 to 32704)  
Saunders,D.C. and Harris,D.  
Unpublished  
3 (bases 1 to 32704)  
Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.  
Direct Submission  
Submitted (03-AUG-2000) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

## COMMENT

Streptomyces coelicolor sequencing at The Sanger Centre is funded  
by the BASRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
CDS are numbered using the following system eg SC7B7.01c. SC (S.  
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary  
strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons  
using a specially developed Hidden Markov Model (Krogh et al.,  
Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot  
program of Bibb et al., Gene 30:157-66(1984) as implemented at  
<http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the  
correct initiation codon. Where possible we choose an initiation  
codon (atg, gtg, ttg or (att)) which is preceded by an upstream  
ribosome binding site sequence (optimally 5-13bp before the  
initiation codon). If this cannot be identified we choose the most  
upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions.  
Cosmid 8A11.

## FEATURES

source  
1..32704  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 8A11"  
misc\_feature  
1..100  
/gene="SC8A11.01"  
/note="nominal overlap with Streptomyces coelicolor cosmid  
ST9A4"  
1..2254  
/gene="SC8A11.01"  
1..2254  
/gene="SC8A11.01"  
/note="SC8A11.01, possible transcriptional regulator  
(fragment), len: >750 aa; similar to TR:Q9S1Z4  
(EMBL:AL109747) Streptomyces coelicolor probable  
transcriptional regulator SCJ21.13, 919 aa; fasta scores:  
opt: 818 z-score: 913.7 E(): 0; 32.6% identity in 775 aa  
overlap. Contains Pfam match to entry PF00196 GerE,  
Bacterial regulatory proteins, luxR family and match to  
Prosite entry PS00622 Bacterial regulatory proteins, luxR  
family signature. Also contains a possible  
helix-turn-helix motif at residues 699..720 (+3.25 SD)"  
/codon\_start=2  
/transl\_table=11  
/product="putative transcriptional regulator (fragment)"  
/protein\_id="CAC01573.1"

/db\_xref="GI:9716212"  
/translation="IGAARTSSPGTYTTLGTHAEVLVLEPLSPSAAAEILDADAGLADA  
VRARILQRAAGNPLALVELPRAAQGISPPDLDLPLTORLETAFAASRTDSLTRECRFTFL  
LVLAAPTAPLNOLLVDVASRLACSEVTYVGLQEAADAGLVLTGRTPEFRHPLARSAI  
YTRATVADRLSTHRALETLEGSPGRLLVHLAATLGPDDELACQLERFADDAKRGQ  
LAAAVPALQAGELVHDPRROTGLLVRAAEALASEINDRVQAQLLNLRDAEAPGTER  
ARLMLVSKAAFEPEPQRIQMDIAAGAFDVGSTVAENLEKRAAARCFQDGA  
RVRAAAALDRWKPDPDPAHVLTVRAYTTPYRRGTOLIAELEKLRDPRDGRLLHYL  
GSGMATGDVGRATRYLAQAASVWRSOGRGLLARSAGSWPRLYLGOLAQAARESAE  
GIALAETGEHWLWGLKATSAITAVLRCGERAARSVRELRAHLSLPVFPVPSYMAQ  
QVGLLALFDSRAVEAYDALARPATKDTDPHHTSRWLLVPDLVDAAGAGRNQARE  
LLVELPELADRLPSMMIVARTYSTAVLADPDHDDTAEDCYDSALSALPDTWPLARLHL  
QHGRRLRQRNRVNDRAKRLARLDEFDRVGAQPMARQELRAAGSDGRRRPSKE  
SLTVQEROIAEGLASQSLNREIGORLFISHTVGAHLRYIVPRLGITSRGRLSAALAA  
LGDDQPTSGC"  
2042..2239  
/gene="SC8A11.01"  
/note="Pfam match to entry PF00196 GerE, Bacterial  
regulatory proteins, luxR family, score 49.10, E-value  
9.7e-11"  
misc\_feature  
2093..2176  
/gene="SC8A11.01"  
/note="PS00622 Bacterial regulatory proteins, luxR family  
signature"  
complement(2184..2453)  
/gene="SC8A11.02c"  
complement(2184..2453)  
/gene="SC8A11.02c"  
/note="SC8A11.02c, unknown, len: 89 aa"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC8A11.02c"  
/protein\_id="CAC01574.1"  
/db\_xref="GI:9716213"  
/translation="MASAATAGTAGALPVMSGOAHADTRGLPSLDWDVDRSVVNPENA  
YTVTVDOVRAEVAQLMGATEETFRPSAGRLRAPRAPRACRGW"  
2787..3257  
/gene="SC8A11.03"  
2787..3257  
/gene="SC8A11.03"  
/note="SC8A11.03, possible regulator, len: 156 aa; similar  
to TR:Q9S2F7 (EMBL:AL096852) Streptomyces coelicolor  
putative regulator SCE19A.24, 142 aa; fasta scores: opt:  
272 z-score: 353.0 E(): 3.4e-12; 36.9% identity in 130 aa  
overlap"  
/codon\_start=1  
/transl\_table=11  
/product="putative regulator"  
/protein\_id="CAC01575.1"  
/db\_xref="GI:9716214"  
/translation="MSGDHGVQAHSAQAALLPLSLCSQMTGALEWEDVPAREFRVD  
PDHPLLVTIRFAPEGAPPVTHVGRDLLHEGLTTSGLGDVQVWADPTDRETAHLQV  
NANGDAIFSLPVELEEDIRTYLHVPAETSSRLTGDAFLSKLDFEPEASSR"  
complement(3297..7058)  
/gene="SC8A11.04c"  
complement(3297..7058)  
/gene="SC8A11.04c"  
/note="SC8A11.04c, probable secreted peptidase, len: 1253  
aa; similar to TR:Q9RL54 (EMBL:AL121596) Streptomyces  
coelicolor probable secreted peptidase SCF31A.10, 1245 aa;  
fasta scores: opt: 1827 z-score: 1848.3 E(): 0; 37.2%  
identity in 1300 aa overlap, to TR:P95684 (EMBL:D83672)  
Streptomyces albobroceolus subtilisin-like protease, 1102  
aa; fasta scores: opt: 1365 z-score: 1380.6 E(): 0; 33.4%  
identity in 1268 aa overlap and to Streptomyces coelicolor  
probable secreted peptidase SC8A11.16c, 1239 aa; fasta  
scores: opt: 1835 z-score: 1467.4 E(): 0; 46.9% identity  
in 1279 aa overlap. Contains Pfam matches to entries  
PF00082 Peptidase\_S8, Subtilase family and PF02225 PA, PA  
domain and matches to Prosite entries PS00136 Serine  
proteases, subtilase family, aspartic acid active site,  
PS00137 Serine proteases, subtilase family, histidine  
active site and PS00138 Serine proteases, subtilase





Db 6006 CCAGGCCGCCGGAGCTGGTGCCGGTGGGTCTCAGCACCTCCAGCTGGACCTGGACCA 5947

QY 354 cagcctggaggccgcgctcggcaagatc 381

Db 5946 CGACCTGGCGGAGCTGATGCGCGGAGC 5919

Search completed: July 18, 2002, 11:22:48  
Job time: 13778 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:37:34 ; Search time 646.57 Seconds  
(without alignments)  
1080.756 Million cell updates/sec

Title: US-09-749-185-8

Perfect score: 407

Sequence: 1 atgagcttcctcgtctccga.....gaggagcagaacgcggctg 407

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	407	100.0	407	21	AAZ49731 Streptomyces netro
2	298.2	73.3	407	21	AAZ49728 Streptomyces grise
3	298.2	73.3	438	21	AAZ49727 Streptomyces grise
4	248.6	61.1	407	21	AAZ49730 Streptomyces golde
5	224.6	55.2	407	21	AAZ49729 Streptomyces albus
6	70.8	17.4	45624	22	AAF88315 S. spinosa DNA fra
7	70.8	17.4	50000	22	AAF88312 S. spinosa DNA fra
8	56.4	13.9	1227	22	AAD14511 Streptomyces clavu
9	56.4	13.9	15079	22	AAD14499 Streptomyces clavu

10	56	13.8	38734	20	AAZ32020 Human METH1 relate
11	56	13.8	38734	22	AAZ90077 AL021529 cDNA clon
12	55.8	13.7	109519	22	AAS08693 Micromonospora DNA
13	54.8	13.5	15079	16	AAQ91580 S. clavuligerus cl
14	51.8	12.7	3241	24	AAS18442 Contig 93 DNA enco
15	51.8	12.7	65140	22	AAAD17184 Streptomyces nous
16	51.8	12.7	125401	22	AAAD17186 Streptomyces nous
17	50.4	12.4	3957	22	AAA09686 HSV-2 immediate ea
18	50.4	12.4	80161	20	AAZ21501 DNA fragment of Sa
19	50.4	12.4	154746	24	AAD25519 Human herpesvirus
20	50.4	12.4	154746	24	AAD25519 Human herpesvirus
21	50	12.3	1761	21	AAA58473 Nucleotide sequenc
22	49.8	12.2	2791	16	AAQ99776 Phosphomycin biosy
23	49.2	12.1	58857	21	AAA58471 Nucleotide sequenc
24	49	12.0	2249	18	AAT74283 Cellulose binding
25	49	12.0	2279	18	AAT74281 Cellulose binding
26	49	12.0	2300	18	AAT74282 Cellulose binding
27	49	12.0	3187	17	AAT10922 Laccase gene. Myc
28	49	12.0	3192	18	AAT63318 Myceliophthora the
29	49	12.0	3192	21	AAZ24236 M. thermophila lac
30	49	12.0	4213	22	AAD06345 Sorghum dwarfing g
31	49	12.0	6827	22	AAD06344 Consensus sequence
32	48.8	12.0	1161	15	AAQ63295 Pseudomonas aerugi
33	48.8	12.0	3954	23	AAAS1481 S. clavuligerus cl
34	48.6	11.9	978	22	AAD14506 P. denitrificans ge
35	48.6	11.9	3855	12	AAQ13287 Streptomyces clavu
36	48.6	11.9	11604	22	AAAD14501 eryA region of S.
37	48.4	11.9	29879	14	AAQ48806 Platenolide syntha
38	48	11.8	44377	18	AAT80414 Pseudomonas aerugi
39	48	11.8	44377	18	AAZ54344 smmR coding sequen
40	47.8	11.7	1425	23	AAQ35141 Bovine herpesvirus
41	47.8	11.7	2312	14	AAQ35141 Bovine herpesvirus
42	47.6	11.7	2799	19	AAV49431 Bovine herpesvirus
43	47.6	11.7	3382	16	AAQ48249 BHV1 gi glycoprote
44	47.6	11.7	3519	16	AAQ94352 Myceliophthora the
45	47.4	11.6	3192	18	AAT72106

#### ALIGNMENTS

RESULT	1
AAZ49731	
ID	AAZ49731 standard; DNA; 407 BP.
XX	AAZ49731;
AC	AAZ49731;
XX	
DT	18-APR-2000 (first entry)
XX	
DE	Streptomyces netropsis ssgA gene.
XX	
KW	ssgA; liquid culture; filamentous bacteria; secondary metabolite;
KW	mycelium; antibiotic, antitumor agent; immunosuppressive agent;
KW	hypocholsterolaeic agent; enzyme inhibitor; antimigraine agent;
KW	herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23940;
KW	bioinsecticide; receptor agonist; antagonist; biomass; ds.
XX	
OS	Streptomyces netropsis.
XX	
FH	Key
FT	mat_peptide
FT	Location/Qualifiers
FT	1..405
FT	/*tag= a
FT	/label= SsgA_protein
XX	
PN	WO200000613-A1.
XX	
PD	06-JAN-2000.
XX	
PF	25-JUN-1999; 99WO-NL00395.
XX	
PR	26-JUN-1998; 98EP-0202148.
XX	
PA	(UYLE-) RIJKSUNIV LEIDEN.

```
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.
XX
XX Van Wezel GP, Kraal B, Luiten RGM;
XX
XX WPI; 2000-147269/13.
XX P-PSDB; AAY44652.
XX
XX Reducing branching and enhancing fragmentation in filamentous
XX microorganisms used to improve their liquid culturing properties -
XX
XX Disclosure; Fig 5; 60pp; English.
XX
XX The present sequence is a DNA (ATCC 23940) encoding S. netropsis SsgA
XX protein. SsgA reduces branching and fragment septation and enhances
XX fragmentation of mycelium in liquid culture resulting in lower viscosity
XX of culture broths. Filamentous bacteria can be transformed with ssgA
XX gene-containing plasmid to enhance the production of secondary
XX metabolites such as, antibiotics, antitumour agents, immunosuppressive
XX agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine
XX agents, herbicides, antiparasitic agents, ruminant growth promoters,
XX bioinsecticides, receptor agonists and antagonists and biomass.
XX
XX Sequence 407 BP; 63 A; 151 C; 131 G; 62 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 407; DB 21; Length 407;
XX Best Local Similarity 100.0%; Pred. No. 6.9e-67;
XX Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 atgagctctctgctccgagagctctctcctcaagatccacgactcgatacagag 60
XX Db 1 atgagctctctgctccgagagctctctcctcaagatccacgactcgatacagag 60
XX
XX Qy 61 accgaggatccctacgaggtcgagatgacattccacatcccccggagagcgctgtgacc 120
XX Db 61 accgaggatccctacgaggtcgagatgacattccacatcccccggagagcgctgtgacc 120
XX
XX Qy 121 tgggcggttcggcgaggagctgctgctgacgggagatcaaccggcgagcgagcgagcgac 180
XX Db 121 tgggcggttcggcgaggagctgctgctgacgggagatcaaccggcgagcgagcgagcgac 180
XX
XX Qy 181 gtccacatcgcccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 240
XX Db 181 gtccacatcgcccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 240
XX
XX Qy 241 gggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 300
XX Db 241 gggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 300
XX
XX Qy 301 acggacaagtgcggtcggtcgaggaacagactctgggtgacttcgaggaagcgctg 360
XX Db 301 acggacaagtgcggtcggtcgaggaacagactctgggtgacttcgaggaagcgctg 360
XX
XX Qy 361 gaggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 407
XX Db 361 gaggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 407
XX
XX RESULT 2
XX AA249728
XX ID AA249728 standard; DNA; 407 BP.
XX
XX AC AA249728;
XX
XX XX 18-APR-2000 (first entry)
XX
XX DE Streptomyces griseus ssgA gene-2.
XX
XX KW ssgA; liquid culture; filamentous bacteria; secondary metabolite;
XX KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;
XX KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;
XX KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23345;
XX KW bioinsecticide; receptor agonist; antagonist; biomass; ds.
XX
```

```
XX Streptomyces griseus.
XX
XX Key Location/Qualifiers
XX mat_peptide 1..405
XX /*tag= a
XX /label= SsgA_protein
XX
XX WO200000613-A1.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-NL00395.
XX
XX 26-JUN-1998; 98EP-0202148.
XX
XX (UYLE-) RIJKSUNIV LEIDEN.
XX (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.
XX
XX Van Wezel GP, Kraal B, Luiten RGM;
XX
XX WPI; 2000-147269/13.
XX P-PSDB; AAY44649.
XX
XX Reducing branching and enhancing fragmentation in filamentous
XX microorganisms used to improve their liquid culturing properties -
XX
XX Disclosure; Fig 5; 60pp; English.
XX
XX The present sequence is a DNA (ATCC 23345) encoding S. griseus SsgA
XX protein. SsgA reduces branching and fragment septation and enhances
XX fragmentation of mycelium in liquid culture resulting in lower viscosity
XX of culture broths. Filamentous bacteria can be transformed with ssgA
XX gene-containing plasmid to enhance the production of secondary
XX metabolites such as, antibiotics, antitumour agents, immunosuppressive
XX agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine
XX agents, herbicides, antiparasitic agents, ruminant growth promoters,
XX bioinsecticides, receptor agonists and antagonists and biomass.
XX
XX Sequence 407 BP; 60 A; 139 C; 142 G; 66 T; 0 other;
XX
XX
XX Query Match 73.3%; Score 298.2; DB 21; Length 407;
XX Best Local Similarity 83.3%; Pred. No. 8.4e-47;
XX Matches 339; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
XX
XX Qy 1 atgagctctctgctccgagagctctctcctcaagatccacgactcgatacagag 60
XX Db 1 atgagctctctgctccgagagctctctcctcaagatccacgactcgatacagag 60
XX
XX Qy 61 accgaggatccctacgaggtcgagatgacattccacatcccccggagagcgctgtgacc 120
XX Db 61 accgaggatccctacgaggtcgagatgacattccacatcccccggagagcgctgtgacc 120
XX
XX Qy 121 tgggcggttcggcgaggagctgctgctgacgggagatcaaccggcgagcgagcgagcgac 180
XX Db 121 tgggcggttcggcgaggagctgctgctgacgggagatcaaccggcgagcgagcgagcgac 180
XX
XX Qy 181 gtccacatcgcccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 240
XX Db 181 gtccacatcgcccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 240
XX
XX Qy 241 gggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 300
XX Db 241 gggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 300
XX
XX Qy 301 acggacaagtgcggtcggtcgaggaacagactctgggtgacttcgaggaagcgctg 360
XX Db 301 acggacaagtgcggtcggtcgaggaacagactctgggtgacttcgaggaagcgctg 360
XX
XX Qy 361 gaggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 407
XX Db 361 gaggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 407
```

```

RESULT 3
ID AAZ49727
XX AAZ49727 standard; DNA; 438 BP.
AC AAZ49727;
XX
DT 18-APR-2000 (first entry)
XX
DE Streptomyces griseus ssgA gene-1.
XX
KW ssgA; liquid culture; filamentous bacteria; secondary metabolite;
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;
KW hypcholesterolaemic agent; enzyme inhibitor; antimigraine agent;
KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 23345;
KW bioinsecticide; receptor agonist; antagonist; biomass; ds.
XX
OS Streptomyces griseus.
XX
FH Key Location/Qualifiers
FT CDS 31..438
FT /tag= a
FT /product= "SsgA protein"
XX
PN WO200000613-A1.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-NL00395.
XX
PR 26-JUN-1998; 98EP-0202148.
XX
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.
XX
PI Van Wezel GP, Kraal B, Luiten RGM;
XX
DR WPI; 2000-147269/13.
DR P-PSDB; AAY44649.
XX
PT Reducing branching and enhancing fragmentation in filamentous
PT microorganisms used to improve their liquid culturing properties -
XX
PS Claim 1; Page 27; 60pp; English.
XX
CC The present sequence is a DNA (ATCC 23345) encoding S. griseus SsgA
CC protein. SsgA reduces branching and fragment septation and enhances
CC fragmentation of mycelium in liquid culture resulting in lower viscosity
CC of culture broths. Filamentous bacteria can be transformed with ssgA
CC gene-containing plasmid to enhance the production of secondary
CC metabolites such as, antibiotics, antitumour agents, immunosuppressive
CC agents, hypcholesterolaemic agents, enzyme inhibitors, antimigraine
CC agents, herbicides, antiparasitic agents, ruminant growth promoters,
CC bioinsecticides, receptor agonists and antagonists and biomass.
XX
SQ Sequence 438 BP; 68 A; 145 C; 153 G; 72 T; 0 other;

Query Match 73.3%; Score 298.2; DB 21; Length 438;
Best Local Similarity 83.3%; Pred. NO. 8.4e-47;
Matches 339; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1 atgagcttcctcgctccgagagctctcctcaagatcccgatccgagcgtacacgac 60
DB 31 atgagcttcctcgctccgagagctctcctcgatcccgatcccgagcgtacacgac 90
QY 61 acccgggatccctacgagcggtcggtgacgtccacccctcccgagagcgctgtgacc 120
DB 91 gtcgagcgtatgccatccgagatgacgtccacccctcccgagcgtacccctgtgacc 150
QY 121 tggcggttcggcgagcggtgctgtcgacgggatacaacccgcccggcgagcgacgac 180

```

```

DB 151 tggcggttcggcgagagctgtgtgtgacgggctcaacagcccgcgagcgacgcat 210
QY 181 gtccacatgcgcccgaccgagcgagggcgtgtcgagctctccatccggtcccagggtg 240
DB 211 gtgcacatgcgcccgaccgagcgagggcctcgagatgtccacatccggtcccagggtc 270
QY 241 ggcgagcgaccgagcctcttcggtgcagcgcccccgcccgctggtccttcctcgaccgc 300
DB 271 ggcgagcgaccgtgctgtccggcgggcgagcgaccgctggtgcttcctcgaccgcg 330
QY 301 acggagacaactcggtccgctcgatcaggaacagactctggtgacttcgagacagccctg 360
DB 331 acggagacaactcggtccgctcgacggcgagcgacacgctgggtgacttcgaggaacacctg 390
QY 361 gaggcgcgctcgagcagatcctccgagggagcagacgagcggcgctg 407
DB 391 gaggcgcaactggcgcgcatcctccgagggagcagacgacgagcgcgctg 437

RESULT 4
ID AAZ49730
XX AAZ49730 standard; DNA; 407 BP.
AC AAZ49730;
XX
DT 18-APR-2000 (first entry)
XX
DE Streptomyces goldeniensis ssgA gene.
XX
KW ssgA; liquid culture; filamentous bacteria; secondary metabolite;
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;
KW hypcholesterolaemic agent; enzyme inhibitor; antimigraine agent;
KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 21386;
KW bioinsecticide; receptor agonist; antagonist; biomass; ds.
XX
OS Streptomyces goldeniensis.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..405
FT /tag= a
FT /label= SsgA_protein
XX
PN WO200000613-A1.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-NL00395.
XX
PR 26-JUN-1998; 98EP-0202148.
XX
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.
XX
PI Van Wezel GP, Kraal B, Luiten RGM;
XX
DR WPI; 2000-147269/13.
DR P-PSDB; AAY44651.
XX
PT Reducing branching and enhancing fragmentation in filamentous
PT microorganisms used to improve their liquid culturing properties -
XX
PS Disclosure; Fig 5; 60pp; English.
XX
CC The present sequence is a DNA (ATCC 21386) encoding S. goldeniensis SsgA
CC protein. SsgA reduces branching and fragment septation and enhances
CC fragmentation of mycelium in liquid culture resulting in lower viscosity
CC of culture broths. Filamentous bacteria can be transformed with ssgA
CC gene-containing plasmid to enhance the production of secondary
CC metabolites such as, antibiotics, antitumour agents, immunosuppressive
CC agents, hypcholesterolaemic agents, enzyme inhibitors, antimigraine
CC agents, herbicides, antiparasitic agents, ruminant growth promoters,
CC bioinsecticides, receptor agonists and antagonists and biomass.
XX

```

```
SQ Sequence 407 BP; 56 A; 135 C; 145 G; 71 T; 0 other;

Query Match      61.18; Score 248.6; DB 21; Length 407;
Best Local Similarity 75.78; Pred. No. 1.2e-37;
Matches 308; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 1 atgagcttcctcgtccgagagctctcctcaagatccacagtcagactcgacgtacagag 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 atgagcttcctcgtccgagagactctcctccgtattccggtggagctgcttacagag 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 acccgggatccctacgcggtgagatgacctccacatcccccagagacgcccgtgtgacc 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 acctgtatccctacgcggtgagatgacctccacatcccccagagatgcccgtgtgacc 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 tggcgcttcggcgagagctctgctgacgggatcaacgcgcccagcgagcgagcgagac 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 tggcgcttcggcgagagctctgctgacgggatgaggttcggtgagggagcgagcgagac 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 181 gtccacatcgcccgacgacccgagggcctgctgacgtctccatcccggtccaggtg 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 gtccacatcgcccgacgacccgagggcctgctgacgtctccatcccggtccaggtg 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 241 ggcgagaccgccccttcctgctgagggcgcccgccgctggtcgcttctctcgacgcg 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 ggcgagaccgagcgatgttcgggtcgacgacgagcgccgctggtcgcttctctcgacgcg 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 301 acgagaaagtgcgctgcgttcgagaaacagactctgggtgacttcgagagacgctg 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 acgagaaagtgcgctgcgttcgagagagagcggttccctcgccacttcgacgctgctc 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 361 gagcgccgctcgcaagatcctcgcgagagagcagacgacgacgacgacgacgacgac 407
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 gagcgccgctcgcaagatcctcgcgagagagcagacgacgacgacgacgacgacgac 407
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AAZ49729
ID AAZ49729 standard; DNA; 407 BP.
XX
AC AAZ49729;
XX
DT 18-APR-2000 (first entry)
XX
DE Streptomyces albus G ssgA gene.
XX
KW ssgA; liquid culture; filamentous bacteria; secondary metabolite;
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;
KW herbicide; antiparasitic agent; ruminant growth promoter; ATCC 3004;
KW bioinsecticide; receptor agonist; antagonist; biomass; ds.
XX
OS Streptomyces albus G.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..405
FT /tag= a
FT /label= SsgA_protein
XX
PN WO200000613-A1.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-NL00395.
XX
PR 26-JUN-1998; 98EP-0202148.
XX
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.
XX
FI Van Wezel GP, Kraal B, Luiten RGM;
XX WPI; 2000-147269/13.
XX
```

```
DR P-PSDB; AAY44650.
XX
PT Reducing branching and enhancing fragmentation in filamentous
PT microorganisms used to improve their liquid culturing properties
XX
PS Disclosure; Fig 5; 60pp; English.
XX
CC The present sequence is a DNA (ATCC 3004) encoding S. albus G SsgA
CC protein. SsgA reduces branching and fragment septation and enhances
CC fragmentation of mycelium in liquid culture resulting in lower viscosity
CC of culture broths. Filamentous bacteria can be transformed with ssgA
CC gene-containing plasmid to enhance the production of secondary
CC metabolites such as, antibiotics, antitumour agents, immunosuppressive
CC agents, hypocholesterolaemic agents, enzyme inhibitors, antimigraine
CC agents, herbicides, antiparasitic agents, ruminant growth promoters,
CC bioinsecticides, receptor agonists and antagonists and biomass.
XX
SQ Sequence 407 BP; 57 A; 150 C; 139 G; 61 T; 0 other;

Query Match      55.28; Score 224.6; DB 21; Length 407;
Best Local Similarity 72.08; Pred. No. 3.3e-33;
Matches 293; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 1 atgagcttcctcgtctccgagagactctcctcaagatccacagtcagactcgacgtacag 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 atgagcttcctcgtctccgagagactctcctccgactcccggtggagctgctgtacag 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 acccgggatccctacgcggtgctgctgacgtacacctccacatcccccagagacgcccgtgtgacc 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 accgtgatccgtacgcggtgctgctgacgtacacctccacatcccccagagacgcccgtgtgacc 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 tggcgcttcggcgagagctgctgctgacgggatcaacgcgcccagcgagcgagcgagcgagac 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 tgggtctctggcggtgactgctggtgagggaggtctctgacgcccggcgagcgagcgagac 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 181 gtccacatcgcccgacgacccgagggcctgctgacgtctccatcccggtccaggtg 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 gtccggtgtcgccggtggggcagacgcccacagggaggtgacatcacctccaggtc 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 241 ggcgagaccgcgccccttcctgctgagggcgcccgccgctggtcgcttctctcgacgcg 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 ggcctccgagcagcgctcttcctgctgagggcgcccgccgctgctgcttctctcgacgcg 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 301 acgagaaagtgcgctgcgttcgagaaacagactctgggtgacttcgagagacgctg 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 accgaccagggctgtgctgcgtcgagcgagcgggcgagcgagcgagcgagcgagcgagac 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 361 gagcgccgctcgcaagatcctcgcgagagagcagacgacgacgacgacgacgacgacgac 407
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 gagcgctctgaaccgacgacctcgcgagagagcagagcgagcgagcgagcgagcgagcg 407
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
AAF88315/c
ID AAF88315 standard; DNA; 45624 BP.
XX
AC AAF88315;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment SEQ ID 4.
XX
KW Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
```

PF 29-NOV-1999; 99DE-1057268.  
PR 27-AUG-1999; 99DE-1040596.  
XX (FARB ) BAYER AG.  
FA Eberz G, Moehrie V, Froede R, Velten R, Salas JA;  
PI WPI; 2001-267102/28.  
XX  
DR New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for  
PT recombinant production of insecticidal spinosyns and their derivatives  
PT  
PT  
PS Claim 7; Page 58-74; 354pp; German.  
XX  
CC This invention describes a novel method nucleic acid (I) and its encoded  
CC polypeptide (II) containing at least one region that encodes an enzymatic  
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to  
CC identify, inactivate or modulate genes involved in the biosynthesis of  
CC (II); (ii) to generate a library of polyketide synthases; (iii) for  
CC adding forosamine or trimethylrhannose to a spinosyn or polyketide  
CC aglycone; and (iv) for recombinant production of the corresponding  
CC enzymes, which are used for production of (II), their precursors or  
CC derivatives, including production of transgenic plants that express (II)  
CC and thus have increased resistance to insects. (I) are also useful as  
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are  
CC also used to raise specific antibodies, useful for identifying  
CC expression clones in a gene bank. Cells transformed with (I) may produce  
CC (II) at significantly increased levels or produce new derivatives of  
CC (II). This sequence represents a genomic DNA fragment of the S. spinosa  
CC genome which contains the coding regions for proteins involved in  
CC forosamine and trimethylrhannose biosynthesis.  
XX  
SQ Sequence 45624 BP; 7933 A; 14913 C; 14940 G; 7838 T; 0 other;

Query Match 17.4%; Score 70.8; DB 22; Length 45624;  
Best Local Similarity 52.3%; Pred. No. 6.3e-05;  
Matches 181; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

QY 38 tccagtcgaactcgatcacgagaccggggtccctacgcgtgctgagatccctccacc 97  
DB 21036 TCGGGGTGGAGTGGGTACGACACGCCGCAATCCGTACGATCCATGAAGCTCAACG 20977  
QY 98 tcccggagagcgctgtgacctggggtgctggcgggagctgtctgtcagcgggata 157  
DB 20976 TAGGCACGGAGCGGTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 20917  
QY 158 accgcccagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 217  
DB 20916 TCGCCGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 20860  
QY 218 acgttcctccatcgctccaggtggtggtggtggtggtggtggtggtggtggtggt 277  
DB 20859 TGGTCGTGATCGAGATGAGTGCCTGCGGGGAGCGGAGCGGAGCGGAGCGGAGCG 20800  
QY 278 cgtgtgtgcttccttcagcagcagcagcagcagcagcagcagcagcagcagcagc 337  
DB 20799 AGCTTGGCGGAGCTTCTTGAACGACACCTACGAGGTGGTGGTGGTGGTGGTGGT 20740  
QY 338 tgggtgacttcagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 383  
DB 20739 GGATGAACGTGACGAGGTGCTGACGACGAGTCTCTCGCACAACT 20694

RESULT 7  
AAF88312  
ID AAF88312 standard; DNA; 50000 BP.  
XX  
AC AAF88312;  
XX

DT 28-AUG-2001 (first entry)  
XX S. spinosa DNA fragment SEQ ID 1.  
XX  
XX Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;  
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
KW macrolide; insecticidal; ds.  
XX  
OS Saccharopolyspora spinosa.  
XX  
PN DE19957268-A1.  
XX  
PD 08-MAR-2001.  
XX  
PF 29-NOV-1999; 99DE-1057268.  
XX  
PR 27-AUG-1999; 99DE-1040596.  
XX  
FA (FARB ) BAYER AG.  
XX  
PI Eberz G, Moehrie V, Froede R, Velten R, Salas JA;  
XX WPI; 2001-267102/28.  
DR  
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for  
PT recombinant production of insecticidal spinosyns and their derivatives  
PT  
PT  
PS Claim 7; Page 14-31; 354pp; German.  
XX  
CC This invention describes a novel method nucleic acid (I) and its encoded  
CC polypeptide (II) containing at least one region that encodes an enzymatic  
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to  
CC identify, inactivate or modulate genes involved in the biosynthesis of  
CC (II); (ii) to generate a library of polyketide synthases; (iii) for  
CC adding forosamine or trimethylrhannose to a spinosyn or polyketide  
CC aglycone; and (iv) for recombinant production of the corresponding  
CC enzymes, which are used for production of (II), their precursors or  
CC derivatives, including production of transgenic plants that express (II)  
CC and thus have increased resistance to insects. (I) are also useful as  
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are  
CC also used to raise specific antibodies, useful for identifying  
CC expression clones in a gene bank. Cells transformed with (I) may produce  
CC (II) at significantly increased levels or produce new derivatives of  
CC (II). This sequence represents a genomic DNA fragment of the S. spinosa  
CC genome which contains the coding regions for proteins involved in  
CC forosamine, trimethylrhannose and polyketide synthase biosynthesis.  
XX  
SQ Sequence 50000 BP; 8484 A; 16290 C; 16535 G; 8691 T; 0 other;

Query Match 17.4%; Score 70.8; DB 22; Length 50000;  
Best Local Similarity 52.3%; Pred. No. 6.3e-05;  
Matches 181; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

QY 38 tccagtcgaactcgatcacgagaccggggtccctacgcgtgctgagatccctccacc 97  
DB 24589 tcgggtggtgagctcggttacgacagccgcaatccgtagagatccatgaagctcaacg 24648  
QY 98 tcccggagagcgctgtgacctggggtgctggcgggagctgtctgtcagcgggata 157  
DB 24649 taggcagcagcggtcaggtggtggtggtggtggtggtggtggtggtggtggt 24708  
QY 158 accgcccagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 217  
DB 24709 tcgcccagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 24765  
QY 218 acgttcctccatcgctccaggtggtggtggtggtggtggtggtggtggtggtggt 277  
DB 24766 tgggtgacttcagagcagcagcagcagcagcagcagcagcagcagcagcagc 24825  
QY 278 cgtgtgtgcttccttcagcagcagcagcagcagcagcagcagcagcagcagcagc 337

```

Db 24826 agcttgaggactcttgaacgacacctacgctggtgcgaacctggtgatgaacacgggt 24885
Qy 338 tgggtgacttcgagacagcctgagcggcgctcgccgcaagaatcct 383
Db 24886 ggaatgaacgtcgacgagtgctgagccagctgctctcgacaaacct 24931

RESULT 8
AAD14511
ID AAD14511 standard; DNA; 1227 BP.
AC AAD14511;
XX
XX 01-NOV-2001 (first entry)
XX
XX Streptomyces clavuligerus ORF10 DNA downstream to pcbC gene.
XX
XX Clavulanic acid biosynthesis enzyme; antibiotic; infectious disease;
KW broad spectrum beta-lactamase inhibitor; open reading frame; ORF;
KW pcbC gene; ds.
XX
XX Streptomyces clavuligerus.
XX
XX Key Location/Qualifiers
CDS 1..1227
FT /*tag= a
FT /product= "Open reading frame (ORF)10 protein"
FT /transl_except= (pos:1222..1224, aa:Trp-Glx)
XX
XX US232106-B1.
XX
XX 15-MAY-2001.
XX
XX 30-AUG-1999; 99US-0385028.
XX
XX 29-JAN-1997; 97US-0790462.
XX
XX 08-OCT-1993; 93US-0134018.
XX
XX 06-DEC-1995; 95US-0567801.
XX
XX (UYAL-) UNIV ALBERTA.
XX
XX Jensen SE, Aidoo KA, Paradkar AS;
XX
XX WPI; 2001-342772/36.
XX
XX P-PSDB; AAE07915.
XX
XX Novel enzyme required for clavulanic acid biosynthesis which is useful
XX at broad spectrum beta-lactamase inhibitor -
XX
XX Disclosure; Column 75-76; 75pp; English.
XX
XX The invention relates to DNA sequences encoding enzymes required
XX for clavulanic acid biosynthesis. Clavulanic acid is a broad spectrum
XX beta-lactamase inhibitor and is an important antibiotic for the treatment
XX of infectious diseases. Also provided in the patent is a 15 Kb genomic
XX DNA fragment downstream to pcbC gene from Streptomyces clavuligerus.
XX The genomic DNA comprises 10 open reading frames (ORFs), eight of which
XX are involved in clavulanic acid biosynthesis. The present sequence is
XX ORF10 DNA from S. clavuligerus genomic DNA.
XX
XX Sequence 1227 BP; 161 A; 439 C; 451 G; 176 T; 0 other;

Query Match 13.98; Score 56.4; DB 22; Length 1227;
Best Local Similarity 48.28; Pred. No. 0.035;
Matches 159; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

Qy 75 cgcggtgcggtacacctcccccagagacgcgcctgtgacctggcggttcgscg 134
Db 849 cgaggactgctgcgttcactccatcctgacagacggcggtgcgcggtgga 908
Qy 135 ggaactgctgcacgggatcaaccgcccgcgagcggcgacgtccacatcgccc 194

```

```

Db 909 ggaactcagctcagcatgtgctcatccggcgaggcggtgctgcgtcgtgc 968
Qy 195 gaccgaccccgaggcctgtcgcagctctccatccgcctcaggtggcgacccgcgc 254
Db 969 ggcgggcaaccgggacgagacggtcttcccgcacccgacccgggtgacgtggaccgcga 1028
Qy 255 cctcttcgtgacggcgcccgccgctgctccttcctgcacgcgcacggacaagtgcgt 314
Db 1029 cgcccgcccatctccttccttcgacgcgcacgcacgtgcctggccagtggctggc 1088
Qy 315 gccgctcgttcaggaacagactcgtggtgacttcgagagacagcctggagccgcgtcgg 374
Db 1089 ccgggtgagctgagagatcctcgcgcgctgctgcgtcggtgacgtgccggct 1148
Qy 375 caagatcctcgccgagagagacagacgcgg 404
Db 1149 cgcggtcccttcgagagctgacttcgg 1178

RESULT 9
AAD14499
ID AAD14499 standard; DNA; 15079 BP.
XX
XX AAD14499;
XX
XX 01-NOV-2001 (first entry)
XX
XX Streptomyces clavuligerus 15 Kb genomic DNA fragment.
XX
XX Clavulanic acid biosynthesis enzyme; antibiotic; infectious disease;
KW broad spectrum beta-lactamase inhibitor; ds.
XX
XX Streptomyces clavuligerus.
XX
XX Key Location/Qualifiers
CDS complement (109..1764)
FT /*tag= a
FT /product= "Open reading frame (ORF)11 protein"
FT /transl_except= (pos:112..114, aa:Gly-Glx)
FT misc_feature 1765..2215
FT /*tag= b
FT /note= "Intergenic region"
FT CDS 2216..3937
FT /*tag= c
FT /product= "Clavulanic acid biosynthesis enzyme"
FT /transl_except= (pos:3932..3934, aa:Leu-Glx)
FT misc_feature 3938..3939
FT /*tag= d
FT /note= "Intergenic region"
FT CDS 3940..5481
FT /*tag= e
FT /product= "Clavulanic acid biosynthesis enzyme"
FT /transl_except= (pos:5476..5478, aa:Ala-Glx)
FT misc_feature 5482..5653
FT /*tag= f
FT /note= "Intergenic region"
FT CDS 5654..6595
FT /*tag= g
FT /product= "Clavulanic acid biosynthesis enzyme, CIA"
FT /transl_except= (pos:6590..6592, aa:Leu-Glx)
FT /note= "The CDS corresponds to ORF4 and does not include
FT start codon"
FT misc_feature 6596..6610
FT /*tag= h
FT /note= "Intergenic region"
FT CDS 6611..7588
FT /*tag= i
FT /product= "Clavaminic synthase isozyme (cs2)"
FT /transl_except= (pos:7583..7585, aa:Arg-Glx)
FT

```









```
FT      CDS      /tag= aq
FT      /product= "EvrO"
FT      complement (36235..36963)
FT      /tag= ar
FT      /product= "EvrP"
FT      complement (36998..38026)
FT      /tag= as
FT      /product= "EvrQ"
FT      complement (38072..38566)
FT      /tag= at
FT      /product= "EvrR"
FT      complement (38892..40163)
FT      /tag= au
FT      /product= "EvrS"
FT      complement (40216..40890)
FT      /tag= av
FT      /product= "EvrT"
FT      complement (40999..40902)
FT      /tag= aw
FT      complement (40887..41576)
FT      /tag= ax
FT      /product= "EvrU"
FT      complement (41679..42707)
FT      /tag= ay
FT      /product= "EvrV"
FT      complement (42714..42717)
FT      /tag= az
FT      complement (42810..43799)
FT      /tag= ba
FT      /product= "EvrW"
FT      complement (43807..43811)
FT      /tag= bb
FT      complement (43799..44866)
FT      /tag= bc
FT      /product= "EvrX"
FT      complement (45014..45760)
FT      /tag= bd
FT      /product= "EvrY"
FT      complement (45767..45770)
FT      /tag= be
FT      complement (45962..46714)
FT      /tag= bf
FT      /product= "EvrZ"
FT      complement (45952..45956)
FT      /tag= bg
FT      complement (47156..49234)
FT      /tag= bh
FT      /product= "EvsA"
FT      complement (51627..52715)
FT      /tag= bi
FT      /product= "EvsB"
FT      complement (51629..51622)
FT      /tag= bj
FT      complement (52889..53557)
FT      /tag= bk
FT      /product= "EvsC"
FT      complement (53554..54207)
FT      /tag= bl
FT      /product= "EvsA"
FT      complement (54362..55117)
FT      /tag= bm
FT      /product= "EvsB"
FT      complement (55125..55128)
FT      /tag= bn
FT      complement (55135..56094)
FT      /tag= bo
FT      /product= "EvsC"
FT      complement (56100..56103)
FT      /tag= bp
FT      complement (56184..56813)
FT      /tag= bq
FT      /product= "EvsC2"
FT      complement (56961..58709)
```

```
Query Match      13.7%; Score 55.8; DB 22; Length 109519;
Best Local Similarity 49.5%; Pred. No. 0.035;
Matches 144; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY   85  atgaccttcacatcccgagacgcgctgtgacctggcgcttcggtggcgagctgctg 144
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  94361 atctgcctcccgccgcgggtgctgctggcgtggcctgctgctgcgggtggccctg 94420
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   145  ctcgacgggatcaaacgcccgcgagcgacgagcgagtcacatgcgcccgaccgacccc 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  94421 cgctcgccgcccacgcgcgtgacctgacccgcgcgcgcgcgcgcgcgcgcgcgcgc 94480
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   205  gagggcctgcggagctctccatccggctccaggtggcgagcgacgcgcctcttcct 264
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  94481 gcgctgcgcacgagcgcgtgacctgtccacggcgccgacccgctgcgcgcgcgcgc 94540
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   265  gcaggcgcccgccgctgctgccttcctcgcacgcgcgcgcgcgcgcgcgcgcgcgc 324
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  94541 gcacccgggtacgcgcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 94600
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   325  caggacagactctgggtgacttcaggacagcgcgcgcgcgcgcgcgcgcgcgcgcgc 375
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  94601 cgacggctgcgcgcacgcgcctgcgcgtggacgcgcgcgcgcgcgcgcgcgcgcgc 94651
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AAQ91580
ID  AAQ91580 standard; DNA; 15079 BP.
XX
AC  AAQ91580;
XX
DT  13-NOV-1995 (first entry)
XX
DE  S. clavuligerus cla gene region.
XX
KW  Clavulinic acid; clavulinate; antibiotic; beta-lactamase-inhibitor;
KW  cla gene; ds.
XX
OS  Streptomyces clavuligerus.
XX
FH  Key
CDS  complement (49..1745)
FT      /tag= a
FT      /label= ORF1
FT      2216..3937
FT      /tag= b
FT      /label= ORF2
FT      3940..5481
FT      /tag= c
FT      /label= ORF3
FT      5654..6595
FT      /tag= d
FT      /label= ORF4
FT      /note= "cla gene"
FT      6611..7588
FT      /tag= e
FT      /label= ORF5
FT      7895..9076
FT      /tag= f
FT      /label= ORF6
FT      9241..10908
FT      /tag= g
FT      /label= ORF7
FT      complement (10998..12296)
FT      /tag= h
FT      /label= ORF8
FT      complement (12662..13365)
FT      /tag= i
FT      /label= ORF9
FT      13769..14995
FT      /tag= j
FT      /label= ORF10
```

```
XX CA2108113-A.
PN 09-APR-1995.
PD 08-OCT-1993; 93CA-2108113.
PF 08-OCT-1993; 93CA-2108113.
PR 08-OCT-1993; 93CA-2108113.
XX (UYAL-) UNIV ALBERTA.
PA Aideo KA, Jensen SE, Paradkar AS;
XX WPI; 1995-207301/28.
DR P-PSDB; AAR77858; AAR77859; AAR77860; AAR77861; AAR77862; AAR77863;
DR AAR77864; AAR77865; AAR77866; AAR77867.
XX Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for
PT biosynthesis of the antibiotic in Streptomyces hosts which do not
PT naturally produce clavulanate
XX Disclosure; Fig.2; 4lpp; English.
PS A library of S. clavuligerus NRRL 5741 genomic DNA fragments (15-22
XX kb) constructed in cosmid pLAFR3 was screened for the cla gene using
CC a probe based on a partial N-terminal sequence from the CLA enzyme.
CC Isolated clone K6L1 included a 15kb fragment having the sequence
CC given in AAQ91580 that included the cla gene (ORF4).
XX Sequence 15079 BP; 2097 A; 5488 C; 5324 G; 2170 T; 0 other;
SQ
Query Match 13.5%; Score 54.8; DB 16; Length 15079;
Best Local Similarity 47.9%; Pred. No. 0.06;
Matches 158; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
Qy 75 cgcggtcgagatgacattccaccctcccgagacgcgctgtgacctggcggttcggcgcg 134
Db 14617 cgaggagctgctgctccaccctccatctgcagaacggctgcccgtccgcggtgga 14676
Qy 135 ggagctgtgtcgacggagataaccgcccgcgagcgacggcgacgtccacatcgcccc 194
Db 14677 ggaagtcagctgcacgagtgtctatcccgcgcgagggcggtggtgtgtgtgtc 14736
Qy 195 gaccgaccccgagcgctgtcgagcttccatcgctccaggtggcgcgacccgcg 254
Db 14737 ggcgggcaaccgggacgagagcgttttcccgacccggagcggtggagcgtggaccgga 14796
Qy 255 cctcttcgtcgagcgcccgcgctggtgtgcttctctcgaccgcaacgacagtcggt 314
Db 14797 cgcccgccgcatctcgcttcggccacgcatgacacgagtcgctggccagtgtgctggc 14856
Qy 315 gccctcgttcaggaacagactctggtgtaacttcgagacagcctggaggccgcgtcgg 374
Db 14857 ccgggtgagctggaggagatcctcgcggtgtgtgctgctggatgcccgggtgcccggct 14916
Qy 375 caagatctcgcgagagcagacccgcg 404
Db 14917 cgcggtgccttcgaggagagtggaattccg 14946
RESULT 14
AAS18442
ID AAS18442 standard; DNA; 3241 BP.
XX AAS18442;
AC AAS18442;
XX 12-MAR-2002 (first entry)
XX Contig 93 DNA encoding S. narbonneis polyketide synthase.
DE Narbonolide polyketide synthase; PKS; narbomycin modification enzyme;
XX erythromycin; rapamycin; tylosin; picromycin; methylmycin; neomethylmycin;
KW
```

```
KW agriculture; ds.
XX Streptomyces narbonneis.
OS US6303767-B1.
XX 16-OCT-2001.
PD 05-NOV-1999; 99US-0434288.
PF 05-NOV-1998; 98US-107093P.
PR 27-MAY-1999; 99US-0320878.
XX (KOSA-) KOSAN BIOSCIENCES INC.
PA Belach MC, McDaniel R;
XX WPI; 2002-065495/09.
DR Nucleic acids encoding narbonolide polyketide synthases from
XX Streptomyces narbonneis, useful for the recombinant production of
PT polyketides, e.g. narbomycin -
XX Claim 1; Column 20-22; 24pp; English.
PS The present invention relates to recombinant DNA vectors (cosmids)
XX that encode for the narbonolide polyketide synthase (PKS) enzyme and
CC various narbomycin modification enzymes from Streptomyces narbonneis.
CC The recombinant DNA vectors can be used to produce recombinant ketide
CC synthases and a variety of different polyketides (e.g. erythromycin,
CC rapamycin, tylosin, narbomycin, picromycin, methylmycin and
CC neomethylmycin) for use in agriculture, medicine and health. The
CC recombinant vectors may be used to produce polyketides in relatively
CC high yields. AAS18432-AAS18443 represent contig DNA sequences that
CC encode for S. narbonneis PKS enzymes.
XX Sequence 3241 BP; 520 A; 1174 C; 1125 G; 422 T; 0 other;
SQ
Query Match 12.7%; Score 51.8; DB 24; Length 3241;
Best Local Similarity 47.4%; Pred. No. 0.24;
Matches 155; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
Qy 44 tcgaactcgatcacgagacccgggacccctacgcggtgcggtgaccttccacctcccg 103
Db 1826 tcggtctctctcgactccggtccgcccccccgagcgtgacaaggcggtgcccagg 1885
Qy 104 gagacgcgcctgtgacctggcggttcgccccggagctgctgctgacggatcaaccgccc 163
Db 1886 cgggtgccgcaaggtcgccgagaaacgcgcggtgctctcgcaacgagggccagccc 1945
Qy 164 cgagcgcgagcgagcgagtcacatcgccccgacgaccccgagggcctgtcgagcgtct 223
Db 1946 tgcgctgcggtgacgcgagcgcgacgacgacgacgacgacgacgacgacgacgacg 2005
Qy 224 ccactcgctccagtgggcgcgacccctcttcgtagcgcccgcccgactcgccggtcgtcg 283
Db 2006 caaggtcacggtcctggcgagccccacgtgctccccgactcgccggtcgccgctcg 2065
Qy 284 tcgcttctctcgacgcgacgagcaaaagtgcgtgcgctgcggtgcagaaacagactctgggtg 343
Db 2066 acaccatcaaggccgcggtggtgcgacggtgacgtacgacgaggggtgagggaga 2125
Qy 344 acttcgagagacgcttgagccgcgcg 370
Db 2126 ccttcgggacgcggtatcccgcgcgcg 2152
RESULT 15
AAD17184
ID AAD17184 standard; DNA; 65140 BP.
XX AAD17184;
AC AAD17184;
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 11:34:02 ; Search time 165.56 Seconds  
(without alignments)  
603.847 Million cell updates/sec

Title: US-09-749-185-8  
Perfect score: 407  
Sequence: 1 atgagcttctctgtctccga.....gaggagcagaacgcggctg 407

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56.4	13.9	1227	4	US-09-385-028-23
2	56.4	13.9	15079	4	US-09-385-028-1
3	55.2	13.6	2219	3	US-08-510-646B-17
4	51.8	12.7	3241	4	US-09-434-288-11
5	50.4	12.4	1155	2	US-08-387-942C-7
6	50.4	12.4	80161	3	US-09-036-987A-1
7	50.4	12.4	80161	4	US-09-370-700-1
8	49	12.0	2249	3	US-08-814-052-19
9	49	12.0	2279	3	US-08-814-052-17
10	49	12.0	2300	3	US-08-814-052-18
11	49	12.0	3183	2	US-08-939-218A-1
12	49	12.0	3187	5	PCT-US95-06815-1
13	49	12.0	3192	1	US-08-706-037-26
14	49	12.0	3192	1	US-08-940-661A-1
15	49	12.0	3192	2	US-09-083-485-1
16	49	12.0	3192	2	US-09-005-397-26
17	48.6	11.9	978	4	US-09-385-028-18
18	48.6	11.9	11604	4	US-09-385-028-13
19	48.4	11.9	20235	1	US-07-642-734C-3
20	48.4	11.9	20235	3	US-08-439-009A-3
21	48	11.8	44377	2	US-08-804-227C-7
22	48	11.8	44377	2	US-08-804-198-1
23	47.8	11.7	2312	1	US-07-736-178C-1
24	47.6	11.7	3382	2	US-08-682-847-1
25	47.6	11.7	3519	1	US-08-035-558-1
26	47.2	11.6	1155	2	US-08-387-942C-18
27	47.2	11.6	1512	3	US-08-911-853-6

28	47.2	11.6	1512	4	US-09-479-409-6	Sequence 6, Appli
29	47.2	11.6	1512	4	US-09-479-453-6	Sequence 6, Appli
30	47.2	11.6	12588	2	US-08-387-942C-1	Sequence 1, Appli
31	47.2	11.6	17612	3	US-08-911-853-29	Sequence 29, Appli
32	47.2	11.6	17612	4	US-09-479-409-29	Sequence 29, Appli
33	47.2	11.6	17612	4	US-09-479-453-29	Sequence 38, Appli
34	46.8	11.5	1251	4	US-09-105-537-38	Sequence 21, Appli
35	46.8	11.5	5970	3	US-09-320-878-21	Sequence 1, Appli
36	46.8	11.5	23673	4	US-09-773-816-1	Sequence 2, Appli
37	46.6	11.4	942	2	US-08-446-806-2	Sequence 17, Appli
38	46.6	11.4	942	4	US-09-385-028-17	Sequence 5, Appli
39	46.4	11.4	1879	2	US-08-403-852D-5	Sequence 5, Appli
40	46.4	11.4	1879	3	US-08-510-646B-5	Sequence 6, Appli
41	46.4	11.4	1879	4	US-09-231-818-5	Sequence 6, Appli
42	46	11.3	3624	1	US-07-951-715A-6	Sequence 6, Appli
43	46	11.3	3624	2	US-08-459-448A-6	Sequence 6, Appli
44	46	11.3	3624	3	US-08-459-595A-6	Sequence 6, Appli
45	46	11.3	3624	3	US-08-459-504B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-385-028-23  
; Sequence 23, Application US/09385028  
; Patent No. 6232106  
; GENERAL INFORMATION:  
; APPLICANT: Susan E. Jensen  
; APPLICANT: Kwamena A Aidoo  
; APPLICANT: Ashish S. Paradkar  
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
; Patent No. 6232106  
; TITLE OF INVENTION: Acid Biosynthesis  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN,PPLC  
; STREET: The Jenifer Bulding, 400 Seventh Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: US 08/790,462  
APPLICATION NUMBER: 29-JAN-1997  
FILING DATE: 29-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: D. Douglas Price  
REGISTRATION NUMBER: 24,514  
REFERENCE/DOCKET NUMBER: 1418/P57452US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202 638-6666  
TELEFAX: (202) 39305350  
TELEX: RCA-248593 IDEA UR  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1227 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-385-028-23

Query Match 13.9%; Score 56.4; DB 4; Length 1227;

Best Local Similarity 48.2%; Pred. No. 0.003;  
Matches 159; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

Qy 75 ccggtggtgagcttccaccccccagagcgcctgtgacctggcggttcgccc 134  
Db 849 CGAGAGAGTGTGCGCTTCCACTCCATCGTCGAGAACGGGTGGCCCGCGGTGA 908  
Qy 135 ggagctgtctgacgggatacaaccgcccagcggcagcgcagctccacatcgccc 194  
Db 909 GGAGCTCCAGCTCGACGATGTCTCATCCGGCGGGCGGCGGTGTCTGCTGTC 968  
Qy 195 gaccgacccccagggcctgtcgagcgtctccatcccggttcaggtggcgagccgc 254  
Db 969 GCGGGGCAACCGGGACGAGAGCGTCTCCCGACCCGCGGGTGGACGTGGACCGGA 1028  
Qy 255 cctcttcgtgacggcggcggcggcggcggcggcggcggcggcggcggcggcggc 314  
Db 1029 CGCCCGCGCCATCTCGGCTTCGGCCACGGCATGACACGATGCGCTGGCGCCAGTGGC 1088  
Qy 315 ggcgtcgttcaggaacagactctgggtgacttcgagacagcctggagggcggtcgg 374  
Db 1089 CGGGTGGAGCTGGAGAGATCTTCGCGCGGTGCTGCGCTGGATGCCCGGTGCCCGGCT 1148  
Qy 375 caagatcttcgcccggaggagcagaacgcgg 404  
Db 1149 CGCGGTGCCCTTCGAGGAGCTGGACTCCG 1178

## RESULT 2

US-09-385-028-1  
; Sequence 1, Application US/09385028  
; Patent No. 6232106  
; GENERAL INFORMATION:  
; APPLICANT: Susan E. Jensen  
; APPLICANT: Kwamena A. Aidoo  
; APPLICANT: Ashish S. Paradkar  
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
; Patent No. 6232106  
; TITLE OF INVENTION: Acid Biosynthesis  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC  
; STREET: The Jennifer Building, 400 Seventh Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/385,028  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/790,462  
; FILING DATE: 29-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: D. Douglas Price  
; REGISTRATION NUMBER: 24,514  
; REFERENCE/DOCKET NUMBER: 1418/P57452US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 638-6666  
; TELEFAX: (202) 39305350  
; TELEX: RCA 248593 IDEA UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15079 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Streptomyces clavuligerus  
US-09-385-028-1

Query Match 13.9%; Score 56.4; DB 4; Length 15079;  
Best Local Similarity 48.2%; Pred. No. 0.0028;  
Matches 159; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

Qy 75 cgcgtggtgagcttccaccccccagagcgcctgtgacctggcggttcgccc 134  
Db 14617 CGAGGAGCTGTGCGCTTCCACTCCATCGTCGAGAACGGGTGGCCCGCGGTGA 14676  
Qy 135 ggagctgtctgacgggatacaaccgcccagcggcagcgcagctccacatcgccc 194  
Db 14677 GGAGCTCCAGCTCGACGATGTCTCATCCGGCGGGCGGCGGTGTCTGCTGTC 14736  
Qy 195 gaccgacccccagggcctgtcgagcgtctccatcccggttcaggtggcgagccgc 254  
Db 14737 GCGGGCAACCGGGACGAGAGCGTCTTCGCCGACCCGCGGTGGACCGGA 14796  
Qy 255 cctcttcgtgacggcggcggcggcggcggcggcggcggcggcggcggcggcggc 314  
Db 14797 CGCCCGCGCCATCTCGGCTTCGGCCACGGCATGCACGATGCGCTGGCGCCAGTGGC 14856  
Qy 315 ggcgtcgttcaggaacagactctgggtgacttcgagacagcctggagggcggtcgg 374  
Db 14857 CGGGTGGAGCTGGAGGAGATCTTCGCGCGGTGCTGCGCTGGATGCCCGGTGCC 14916  
Qy 375 caagatcttcgcccggaggagcagaacgcgg 404  
Db 14917 CGCGGTGCCCTTCGAGGAGCTGGACTCCG 14946

## RESULT 3

US-08-510-646B-17  
; Sequence 17, Application US/08510646B  
; Patent No. 6077699  
; GENERAL INFORMATION:  
; APPLICANT: Blanc, Veronique  
; APPLICANT: Blanche, Francis  
; APPLICANT: Crouzet, Joel  
; APPLICANT: Jacques, Nathalie  
; APPLICANT: Lacroix, Patricia  
; APPLICANT: Thibaut, Denis  
; APPLICANT: Zagorec, Monique  
; APPLICANT: Debussche, Valerie  
; APPLICANT: De Crecy-Lagard, Valerie  
; TITLE OF INVENTION: Polypeptides Involved In The  
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/510,646B  
; FILING DATE: 03-AUG-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/403,852  
; FILING DATE: 10-MAY-1995

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR 93/00923  
; FILING DATE: 25-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 92/11441  
; FILING DATE: 25-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REFERENCE/DOCKET NUMBER: 03806.0054-01000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 408-4000  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2219 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: s. virginiae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..2219  
; OTHER INFORMATION: /product = "virginiamycin s synthase gene"  
US-08-510-646B-17

Query Match 13.6%; Score 55.2; DB 3; Length 2219;  
Best Local Similarity 50.3%; Pred. No. 0.005;  
Matches 194; Conservative 0; Mismatches 183; Indels 9; Gaps 2;  
  
Qy 12 cgtctcagagagtgctctctcaagatccccagtcgaactcgatcgatagacccggggtacc 71  
Db 41 CTTGTTCGAGGACCCACCGTCGAGGCGCTCGCGGAAACCCCTGGAAAGAGGCCCGCGAGGT 100  
  
Qy 72 ctacgcggtcgatgaccttccacctcccgagacgcgctgacctgagctggcggttcgg 131  
Db 101 CCGCGCCCGCCCTCG 160  
  
Qy 132 ccgggagctgctcagcggggtacacccgagcggcgagcggcgagcgtccacatcgc 191  
Db 161 GCGCTGTGGTTCCTCGACCCCTTGGAGGACCACTCCACTACACATCCCGCTCGC 220  
  
Qy 192 ccgacgaccccgagggcctctcggagctctccatccggtccaggtgggcgcgagccg 251  
Db 221 CTTGCGCTGCGC---GGCGAGCTGGACCGACCGCGCGCTGCAACAGGCGCCTCACCGACCT 277  
  
Qy 252 cgcctcttcgtagcaggcccgcccgctggtgccttctcgcacgcacgcagcagcaagtc 311  
Db 278 GACGACCGCGCCACAAAGCCTTACCGACCGCTTACCGAGCGCGCGCGCGCGCGCGCGCG 337  
  
Qy 312 ggtgcgctcgtcaggaacagactctggtgacttcgaggaacagcc-----tggaagc 365  
Db 338 GCAGCTCTCCGACCGACGAGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 397  
  
Qy 366 cgcctcggcgaagatctcgcgcagg 391  
Db 398 CGGACTCGCGAGATGCTGGCGGAGG 423

RESULT 4  
US-09-434-288-11  
; Sequence 11, Application US/09434288  
; Patent No. 6303767  
; GENERAL INFORMATION:  
; APPLICANT: Betlach C., Melanie  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; TITLE OF INVENTION: CONSTRUCTS THEREFOR

FILE REFERENCE: 30062-20030.00  
; CURRENT APPLICATION NUMBER: US/09/434,288  
; CURRENT FILING DATE: 1999-11-05  
; PRIOR APPLICATION NUMBER: 60/107,093  
; PRIOR FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 3241  
; TYPE: DNA  
; ORGANISM: Streptomyces narbonensis  
US-09-434-288-11

Query Match 12.7%; Score 51.8; DB 4; Length 3241;  
Best Local Similarity 47.4%; Pred. No. 0.022;  
Matches 155; Conservative 0; Mismatches 172; Indels 0; Gaps 0;  
  
Qy 44 tcgaactcgatcacgagacccggggtacccctacgcggtgctgacatccaccttccacctccccc 103  
Db 1826 tcggtctgctcgtcgactccgcccgcgcccccccgagcgtagacaagcgggcgcccagg 1885  
  
Qy 104 gacagcgccctgtgacctggcggttcggccgggagctgctgctgacggggtacacccgcc 163  
Db 1886 cgggtcccgcaaggtcgccgagaacgcgcggtgctcctcgcaacgagggcgccagggccc 1945  
  
Qy 164 cgagcggcgacgagcgtccacatcgccccgacgacccgagggcctgtcgagcgtct 223  
Db 1946 tgcgcgtcgcggtgacgcccggcaagagcgtcgccgctcgcgcccgcgctcgacc 2005  
  
Qy 224 ccattcggctccagtggtggcgcgaccgcccctcttcgtgacgagcggcccgccgctgg 283  
Db 2006 ccaagtcacgctggtggcgagcggccacgctgctccgactcgcgcgcgccgctcg 2065  
  
Qy 284 tcgctcttcgacgcacgacgaagtcggtcgccgctggtcgccgagcagacactctgggtg 343  
Db 2066 acaccataagcggcgcgcggtgagcgtgacgtacgagacgagcgggtgaggaga 2125  
  
Qy 344 acttcgagcagcctggagccgcgc 370  
Db 2126 ccttcggagcggatccccggcggcgc 2152

RESULT 5  
US-08-387-942C-7  
; Sequence 7, Application US/08387942C  
; Patent No. 5939289  
; GENERAL INFORMATION:  
; APPLICANT: ERTESVAG, HELGA  
; APPLICANT: VALLA, SVEIN  
; APPLICANT: SKJAK-BRAEK, GUDMUND  
; APPLICANT: LARSEN, BJORN  
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
; STREET: P.O. BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22042  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/387,942C  
; FILING DATE: 09-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M.

```

; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1809-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-387-942C-7

```

	Query Match	12.4%	Score 50.4	DB 2	Length 1155
	Best Local Similarity	58.8%	Pred. No. 0.042		
	Matches 87	Conservative 0	Mismatches 61	Indels 0	Gaps 0
QY	115	gtgacctgggcgttccggccggga	ctgctcgcagcggggatcaacccgcgcagcggcgac	174	
Db	265	GTGGCGTTCGGCGCTACGGCGAGGAGCACCAGCAACATTCGGCATTCGACGACTGACCTCGAC	324		
QY	175	ggcgacgtccacatcgcgcccgaccgcgcgagggcctgtcgagcgtctccatccgcggtc	234		
Db	325	GGCAACCGGCACACACCCAGCGGGAAGGTTCACGGCTGGTTCAACGGGTACATCCCCGGC	384		
QY	235	caggtggcgcgaggaccgcgcctcttc	262		
Db	385	CAGACGCGCGCCGACCGCGACGTGACCC	412		

RESULT 6  
US-09-036-987A-1  
Sequence 1, Application US/09036987A  
Patent No. 6143526  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.  
APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patti J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
TITLE OF INVENTION: Production  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036,987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)337-4816  
TELEFAX: (317)337-4847  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

```

Query Match	12.4%;	Score 50.4;	DB 3;	Length 80161;
Best Local Similarity	51.4%;	Pred.No. 0.039;		
Matches 142;	Conservative	0;	Mismatches 131;	Indels 3;
Gaps	1;			

  

QY	85	atgaccttccacatccccggagagcgccctgtgacctggcgcttcgagccggagactgctg	144
Db	8	ATCAAGCTCAACCTAGGCACGAGCGTCAAGTGGGTGATCGCCCGGACCTGCTG	67
QY	145	ctcgagcgggataaacgc	204
Db	68	GCCGACGGGCTGATCGCCGAGGACGGCGAAGGCGATGTGGGATCGGCCCTTCGA	124
QY	205	gagggcctgtcgcgacgtctccatcccgcgtccagggtggcgcgcgcgcgcgcctttccgt	264
Db	125	GGTTTTCCGGGGTGGTTCGTGATCGAGATGAGTTCGCCGTCGGGGCAGGCGTCTCTCGAG	184
QY	265	gcagcgccccgcgcgtgtcgccttctcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgt	324
Db	185	GTGAATGCTGACCAAGCTTGGCGGACTTCTTGAACGACACCTACGACGGTGGTGGAACTGGT	244
QY	325	caggaacacactcgggtgacttcgaggacgcctg	360
Db	245	GATGAACACCGGTGGATGAACGTCGACGAGGTGCTG	280

```

RESULT      7
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Blosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

```

	Query Match	12.4%	Score 50.4;	DB 4;	Length 80161;
	Best Local Similarity	51.4%;	Pred. NO. 0.039;		
	Matches 142;	Conservative 0;	Mismatches 131;	Indels 3;	Gaps 1;
QY	85	atgaacttccactccgccgagacgcgcctgtgaacctggcggttcggccggagactctg	144		
Db	8	atgaactcaacgtaggacgcgcgcctgagtggaactggatgcgccgcgacctgctg	67		
QY	145	ctcgacgggataacccgcgcgcgcgcgcgcgcacatgccccgcacgcgacccc	204		
Db	68	gccgacggcctgatccgcgagggcaggcgaagtgcggaatcgcccttcga	--cgg 124		
QY	205	gagggcctgtcgagactctccatcccgctccagtgggcggaacgcgcctcttcctg	264		



Db 125 ggtttccgggtgtgctgatgagatgagtcgctcggtgggagcgctccttcagag 184  
Qy 265 gcaggcgcccgctgctgctctctcgacgcgacgagcaagtgcgtgccgctcggt 324  
Db 185 gtgaatgctaccagcttgcggactctctgaacgacactacgacgtggtcgaaactggt 244  
Qy 325 caggaaacagactcgggtgacttcgagacagcctg 360  
Db 245 gatgaacacgggtgatgaacgctgacgaggtgctg 280

## RESULT 8

US-08-814-052-19  
; Sequence 19, Application US/08814052  
; Patent No. 6015783  
; GENERAL INFORMATION:  
; APPLICANT: von der Osten, Claus  
; APPLICANT: Cherry, Joel R.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Vind, Jesper  
; APPLICANT: Rasmussen, Michael Dolberg  
; TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING  
; TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6015783o No. 6015783disk of No. 6015783th America, Inc.  
; STREET: 405 Lexington Avenue, Suite 6400  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/814,052  
; FILING DATE: 06-MAR-1997  
; CLASSIFICATION: 510  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 4684.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-814-052-19

Query Match 12.0%; Score 49; DB 3; Length 2249;  
Best Local Similarity 53.1%; Pred. No. 0.077;  
Matches 128; Conservative 0; Mismatches 110; Indels 3; Gaps 1;

Qy 60 gaccggatccctacgcgggtgcggtgacgtacacctccaccccgagagcgccgtgtgac 119  
Db 1729 gccggggacggccgctgctgagggggcccaaccctgtgcccggggagctgacgatgct 1788  
Qy 120 ctggcgcttcggcgagctgctgctgacggggtatcaaccgcccagcg---ggagcgg 176  
Db 1789 gccggcgcttgcgggtgggtgctgaccttcggcccgccgacacccggcgctgctgtt 1848  
Qy 177 gcaactccacatgcgcccgaacgaccccgagggcctgtcgagctctccatccggctcca 236  
Db 1849 ccactgccacatcgccctggcagcttctcgggggccctggtgctgcttacctcgagcgccg 1908

Qy 237 ggtggcgcgagaccgcgcctcttcctgacggcgcccgccgctggtgccttctctoga 296  
Db 1909 cgagacactgcggggcgctctcgacgacggccgacggcgacactcgagcgcttctgccc 1968  
Qy 297 c 297  
Db 1969 c 1969

## RESULT 9

US-08-814-052-17  
; Sequence 17, Application US/08814052  
; Patent No. 6015783  
; GENERAL INFORMATION:  
; APPLICANT: von der Osten, Claus  
; APPLICANT: Cherry, Joel R.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Vind, Jesper  
; APPLICANT: Rasmussen, Michael Dolberg  
; TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING  
; TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6015783o No. 6015783disk of No. 6015783th America, Inc.  
; STREET: 405 Lexington Avenue, Suite 6400  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/814,052  
; FILING DATE: 06-MAR-1997  
; CLASSIFICATION: 510  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 4684.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2279 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-814-052-17

Query Match 12.0%; Score 49; DB 3; Length 2279;  
Best Local Similarity 53.1%; Pred. No. 0.077;  
Matches 128; Conservative 0; Mismatches 110; Indels 3; Gaps 1;

Qy 60 gaccggatccctacgcggtgcggtgacgtacacctccaccccgagagcgccgtgtgac 119  
Db 1835 gccggggacggccgctgctgagggggcccaaccctgtgcccggggagctgacgatgct 1894  
Qy 120 ctggcgcttcggcgagctgctgctgacggggtatcaaccgcccagcg---ggagcgg 176  
Db 1895 gccggcgcttgcgggtgggtgctgaccttcggcccgccgacacccggcgctgctgtt 1954  
Qy 177 gcaactccacatgcgcccgaacgaccccgagggcctgtcgagctctccatccggctcca 236  
Db 1955 ccactgccacatcgccctggcagcttctcgggggccctggtgctgcttacctcgagcgccg 2014  
Qy 237 ggtggcgcgagaccgcgcctcttcctgacggcgcccgcccgctggtgcgcttctctoga 296





```
; NAME/KEY: CDS
; LOCATION: join(587..832, 918..995, 1080..1091, 1194..1265,
; LOCATION: 1338..2309, 2457..2525, 2619..3029)
; US-08-706-037-26

Query Match      12.0%; Score 49; DB 1; Length 3192;
Best Local Similarity 53.1%; Pred. No. 0.076;
Matches 128; Conservative 0; Mismatches 110; Indels 3; Gaps 1;

QY 60 gaccggggtacccctacggtgagatgacattccacctccccgagagacgacctgtgac 119
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2693 GGCAGCGGAGCGGGCTGCTGACGGGGGCCAACCCCTGTGCGGGGAGCGTGCATGCT 2752
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 120 ctgggcttcggcgagctgctgcacgggatcaaacgccccgagcg--gcgacgg 176
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2753 GCCGGGCTTCGGGTGGTGTCTCTTCGCGGCGACAAACCCGGCGCTGGCTGT 2812
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 cgagctccacatcccccagaccgaggggctgtcggagctctccatcccggtcca 236
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2813 CCACTGCACATCCCTGGCAGCTCTCGGGGGCTGGGCGTCTACCTCGAGCGCGC 2872
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 ggtgggagcgacgagccctcttcgtgagggcgcccccgctggtcgcttccctga 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2873 CGAGGACCTCGCGGGGGCGCTCTCGGACGCGCGACGACCTCGACCGCCTCTGCGC 2932
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 c 297
Db 2933 C 2933

RESULT 14
US-08-940-661A-1
; Sequence 1, Application US/08940661A
; Patent No. 5795760
; GENERAL INFORMATION:
; APPLICANT: BERKA, Randy Michael
; APPLICANT: BROWN, Stephen H.
; APPLICANT: XU, Feng
; APPLICANT: SCHNEIDER, Palle
; APPLICANT: OXENB LL, Karen M.
; APPLICANT: AASLYNG, Dorrit A.
; TITLE OF INVENTION: PURIFIED MYCELIOPTHORA LACCASES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5795760 No. 5795760disk of No. 5795760th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,661A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/441,146
; FILING DATE: 15-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4184.010-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3192 base pairs
; TYPE: nucleic acid
```

```
; LENGTH: 3192 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(586..831, 917..994, 1079..1090, 1193..1264,
; LOCATION: 1337..2308, 2456..2524, 2618..3028)
; US-08-940-661A-1

Query Match      12.0%; Score 49; DB 1; Length 3192;
Best Local Similarity 53.1%; Pred. No. 0.076;
Matches 128; Conservative 0; Mismatches 110; Indels 3; Gaps 1;

QY 60 gaccggggtacccctacggtgagatgacattccacctccccgagagacgacctgtgac 119
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2692 GGCAGCGGAGCGGGCTGCTGACGGGGGCCAACCCCTGTGCGGGGAGCGTGCATGCT 2751
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 120 ctgggcttcggcgagctgctgcacgggatcaaacgccccgagcg--gcgacgg 176
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2752 GCCGGGCTTCGGGTGGTGTCTCTTCGCGGCGACAAACCCGGCGCTGGCTGT 2811
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 cgagctccacatcccccagaccgaggggctgtcggagctctccatcccggtcca 236
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2812 CCACTGCACATCCCTGGCAGCTCTCGGGGGCTGGGCGTCTACCTCGAGCGCGC 2871
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 ggtgggagcgacgagccctcttcgtgagggcgcccccgctggtcgcttccctga 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2872 CGAGGACCTCGCGGGGGCGCTCTCGGACGCGCGACGCGCGACCTCGACCGCCTCTGCGC 2931
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 c 297
Db 2932 C 2932

RESULT 15
US-09-083-485-1
; Sequence 1, Application US/09083485
; Patent No. 5948121
; GENERAL INFORMATION:
; APPLICANT: Dorrit Aaslyng
; APPLICANT: Sorensen, Niels H.
; TITLE OF INVENTION: Laccases with Improved Dyeing
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5948121o No. 5948121disk of No. 5948121th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,485
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 4639.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3192 base pairs
; TYPE: nucleic acid
```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(586..831, 917..994, 1079..1090, 1193..1264,
; LOCATION: 1337..2308, 2456..2524, 2618..3028)
US-09-083-485-1

Query Match      12.08; Score 49; DB 2; Length 3192;
Best Local Similarity 53.1%; Pred. No. 0.076;
Matches 128; Conservative 0; Mismatches 110; Indels 3; Gaps 1;

Qy 60 qaccggggtaccctacgcggtgcggtgacacctccaccccgagagacgcgcctgtgac 119
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2692 GCGCGGGACCGGGCCCTGCTGACGGGGCCCAACCCCTGTGCGGGGGACGTGACGATGCT 2751
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 120 ctggggttcggcgaggagctgctgctgcgacgggatacaacgcggcgagcg---gcgacgg 176
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2752 GCGGCGTTCGGGTGGGTGGTGTGCTGGCTTCGCGGCCGACAAACCGGCGCTGGCTGTT 2811
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 177 cgagtcacatgccccgacccgagcccgagggcctgtcggagctctccatccggctcca 236
   || | ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2812 CCACTGCCACATCCCTGGCACGTCTCGGGCGGCTGGGGCTGCTTACCTCGAGCGCGC 2871
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 237 ggtgggcgcgagccgcctcttccttcggtgcagcgcccccgcgctggtcgcttcctcga 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2872 CGAGCACTGCGCGGGGCCGCTCTCGGACGCGCGACCGACCTCGACCGCCTCTGCGC 2931
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 297 c 297
   |
Db 2932 C 2932

```

Search completed: July 18, 2002, 11:34:28  
Job time: 11458 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2002, 10:12:53 ; Search time 4821.96 Seconds  
(without alignments)  
1139.217 Million cell updates/sec

Title: US-09-749-185-8  
Perfect score: 407  
Sequence: 1 atgagcttctctgtctccga.....gaggagcagaacgcggctg 407

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61.8	15.2	925	12	CNS0091P
2	60.4	14.8	925	12	CNS0091P
3	56	13.8	684	10	BI717260
c	54.6	13.4	704	10	BG858938
c	54.4	13.4	775	10	BG843065
6	53.8	13.2	935	12	CNS006XK
c	53.4	13.1	816	12	AG075896
8	52.2	12.8	844	12	CNS0052P
9	52.2	12.8	957	12	AG077010
10	51.8	12.7	447	9	AU183257
11	51.4	12.6	770	12	AG054664
12	50.6	12.4	469	10	BG274193
13	50.6	12.4	504	10	BE360790
14	50.6	12.4	1434	12	AG058290
c	50.2	12.3	935	12	CNS006XK
16	49.8	12.2	361	10	BF292743
17	49.8	12.2	516	10	BF259400

c	18					
c	19					
	20					
	21					
c	22					
	23					
	24					
	25					
	26					
	27					
	28					
	29					
	30					
	31					
c	32					
	33					
	34					
	35					
	36					
	37					
	38					
	39					
	40					
c	41					
	42					
c	43					
	44					
c	45					

#### ALIGNMENTS

#### RESULT 1

#### CNS0091P

#### LOCUS

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### COMMENT

CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL053013 GI:4934461  
GSS.  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 925)  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aron Mammoser in Pieter de Jong's Laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers  
1. .925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"





```

/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II"
/notes="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
McDerrott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Rosaldo et al. 1996. Genome Research 6: 791-806."

```

122 a	243 c	171 g	168 t
-------	-------	-------	-------

```

Query Match      13.4%; Score 54.6; DB 10; Length 704;
Best Local Similarity 50.2%; Pred. No. 8.1;
Matches 135; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

```

[illegible]

EST 29-MAY-2001	linear
BG858938	704 bp mRNA
LOCUS	C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION	Chlamydomonas reinhardtii cDNA, mRNA sequence.

KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii.  
ACCESSION BG843065  
VERSION BG843065.1 GI:14224249

Source	Organism
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;	<i>Chlamydomonas reinhardtii</i> .
Chlamydomonadaceae; <i>Chlamydomonas</i> .	<i>Chlamydomonas reinhardtii</i> .

**AUTHORS**  
Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,  
McDermott, J. P., Silflow, C., Stern, D., and Surzycki, R.

**REFERENCE**  
1 (bases 1 to 775)  
Chlamydomonadales: Chlamydomonas.

Cellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2

COMMENT	<p>           Contact: Charles Hauser            DCMB Box 91000            unpublished (2000)         </p>
---------	--

Durham, NC 27708-1000

Fax: 919 613 8177

**FEATURES**

/organism="Chla

/db\_xref="taxon

/strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap  
 II"  
 /note="Vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; This library, constructed by John Davies and Jeffrey  
 McWhermott, combines cDNAs from CC-1690 cells grown to  
 mid-log phase in TAP (acetate-containing) medium in the  
 light, TAP medium in the dark, HS (minimal) medium in  
 ambient levels of CO<sub>2</sub> and HS medium bubbled with 5% CO<sub>2</sub>.  
 PolyA mRNA was purified from each sample, pooled and cDNA  
 synthesized. The cDNA was directionally cloned into lambda  
 Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.  
 pBluescript II SK- plasmids were excised from the lambda  
 Zap clones by superinfection with Exassist (Stratagene)  
 phage. The library was normalized using method 4 described  
 in Bonaldo et al (1996) Genome Research 6: 791-806."  
 133 a 265 c 200 q 177 t  
 BASE COUNT

Query Match	13.4%	Score 54.4;	DB 10;	Length 775;
Best Local Similarity	49.6%;	Pred. No. 8.7;		
Matches 139;	Conservative 0;	Mismatches 141;	Indels 0;	Gaps 0;
QY 126	gttcgcccggagctgctgcacagggatcaaacccgccgagcggcgacggcgacgctcca	185		
Db 688	GGTGGCCACCACACTGTGTGGCCCGGTATCGACATTGAGCGCGCTCAACATTTGTCATCAA	629		
QY 186	catcgccccgacgacccccgagggcctgtcgagcgtctccatcgctccaggtggcgcg	245		
Db 628	CTAGGACATGCCCGAGTCCGACGCAAGTCAAGGGCGAGTCCAAAGCAGCGCAACGGCGC	569		
QY 246	ggaccgcgcgcctctccgtgcaggcgcgccgcgcgtggtcgctctctcgaccgcacgga	305		
Db 568	GGACACCTACTGCACCGGTCGCGCCCGCGGGCCGCTTCGGCACCAAGGCTCTGGCGCAT	509		
QY 306	caagtcggtgcgcctcgttcaggacaacagactctgggtgacttcgagcacagcctggaggc	365		
Db 508	CACGTTCTGTGTCGCGAGGAGACACGGCGGTCTCAACGCCGCTGCAGGACCCGCTTCA	449		
QY 366	cgcgctcggcaagatctctcgcggaggagcagaacgcggcg	405		
Db 448	CGTGGACATCAAGCCCTCGCGGAGAGAGATCGACGCCAGC	409		

RESULT	6	
LOCUS	CNS006XK	935 bp DNA linear
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14N09 of RCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	GSS 03-JUN-1999
ACCESSION	AL066051	
VERSION	AL066051.1	GI:4945019
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 935)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr	
	- web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )	
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department	

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RpCi-98 and was constructed by partial *EcoRI* digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES
  source
    Location/Qualifiers
      1..935
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone_lib="RPCI-98"
        /clone="BACR14N09"

```

BASE COUNT	257 a	170 c	162 g	96 t	250 others
ORIGIN					

Query Match	13.2%	Score 53.8;	DB 12;	Length 935;
Best Local Similarity	30.4%;	Pred. No. 11;		
Matches 94;	Conservative	74;	Mismatches 141;	Indels 0;
				Gaps 0;

[illegible]

RESULT	7
LOCUS	AG075896
DEFINITION	Pan troglodytes DNA, clone: PTB-068024.R, genomic survey sequence.
ACCESSION	AG075896
VERSION	AG075896.1 GI:16627698
KEYWORDS	GSS; GSS (genome survey sequence).
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-068024.R.

ORGANISM  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
1 (sites)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
BAC end sequences of Library PTB  
Unpublished  
2 (bases 1 to 816)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.

**TITLE**  
**JOURNAL**

**Direct Submission**  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: [chirmpes@gsc.riken.go.jp](mailto:chirmpes@gsc.riken.go.jp), URL: <http://hqp.gsc.riken.go.jp/>).





```
source
1. .770
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone_lib="PTB-040615.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      61 a  428 c  226 g  21 t  34 others
ORIGIN

Query Match      12.6%; Score 51.4; DB 12; Length 770;
Best Local Similarity 46.7%; Pred. No. 26;
Matches 148; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 89 ccttcacctcccgagacgcgctgtgacctggcggttcggcggagctgctgcg 148
   || || || || || || || || || || || || || || || || || || ||
Db 396 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 455
   || || || || || || || || || || || || || || || || || || ||
QY 149 acgggatcaaccgcgagcgcgacgcgcacgtccacatcgccccgacgacccgag 208
   || || || || || || || || || || || || || || || || || || ||
Db 456 CCGCACACCCCAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 515
   || || || || || || || || || || || || || || || || || || ||
QY 209 gcctgtgcgagcttcacatcgctccaggtggcgcgacccgcgcctcttcctgcg 268
   || || || || || || || || || || || || || || || || || || ||
Db 516 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 575
   || || || || || || || || || || || || || || || || || || ||
QY 269 gcgcgccgcgcgtggtgccttcctgcacgcgacggaagtgcgtgcgtgcgtag 328
   || || || || || || || || || || || || || || || || || || ||
Db 576 CCGNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 635
   || || || || || || || || || || || || || || || || || || ||
QY 329 aacagactctgggtgacttcgagacagcctgagcgcgctcgcaagatctcgcg 388
   || || || || || || || || || || || || || || || || || || ||
Db 636 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 695
   || || || || || || || || || || || || || || || || || || ||
QY 389 aggagcagaacgcgcgc 405
   || || || || || || || || || || || || || || || || || || ||
Db 696 CCGCGCGCGCGCGCGCG 712

RESULT 12
BG274193
LOCUS
DEFINITION WHE2230_F05_K10ZS Aegilops speltoides anther cDNA library Aegilops
speltoides cDNA clone WHE2230_F05_K10, mRNA sequence.
ACCESSION BG274193
VERSION BG274193.1 GI:13066255
KEYWORDS EST.
SOURCE Aegilops speltoides.
ORGANISM Aegilops speltoides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
1; Triticeae; Aegilops.
REFERENCE 1 (bases 1 to 469)
AUTHORS Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J.,
Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R.,
Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C.
and Zhang,D.
TITLE The structure and function of the expressed portion of the wheat
genomes - Anther cDNA library from Aegilops speltoides
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: andersnpw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratiogene SK primer.
Location/Qualifiers
source
1. .469

/organism="Aegilops speltoides"
/cultivar="F2 from 2-12-4-8-1-1-1-(1) x PI36909-12-811-(1)"
/db_xref="taxon:4573"
/clone="WHE2230_F05_K10"
/clone_lib="Aegilops speltoides anther cDNA library"
/tissue_type="Anther"
/dev_stage="Premeiotic anthers"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth
chamber at the University of California, Davis (Akhunov).
Premeiotic anthers were harvested, total RNA and poly(A)
RNA were prepared, from each tissue and then pooled, a
cDNA library was made, and the cDNA clones were in vivo
excised to give pBluescript phagemids in the TJ Close lab
(Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons
, Zhang) at the University of California, Riverside.
Plasmid DNA preparations and DNA sequencing were performed
in the OD Anderson lab (all other authors)."
```

Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: JEN REV  
High quality sequence stop: 477  
POLYA=No.

FEATURES  
source Location/Qualifiers  
1..504  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4538"  
/clone\_lib="Dark Grown 1 (DGI)"  
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."  
BASE COUNT 91 a 216 c 133 g 64 t  
ORIGIN

Query Match 12.4%; Score 50.6; DB 10; Length 504;  
Best Local Similarity 49.0%; Pred. No. 35;  
Matches 165; Conservative 0; Mismatches 169; Indels 3; Gaps 1;  
QY 52 cgatacagacccgggataccctacgcggtgctgagatgacctccacctcccccggagacgcg 111  
Db 4 CGAGGCCAGCAGCGGACCTCATCCACCACCACTACATCCCGACCTCATCGAGGCGC 63  
QY 112 cctgtgacctggcgctcgccgggagactgctgacagggatcaaccgccagcgcc 171  
Db 64 ACCGGATGCGGTCAAGTCGCGGAGCGACGGCTCTAC---ACCAACCGCGCACCGGG 120  
QY 172 gacggcagctccacatcgccccagccagccagggcctgctgagcgtctccatccg 231  
Db 121 CCGCGCACACACACACACCGCTCTGGACGTGGACGCTTCGACCCGCTCACGTTTC 180  
QY 232 ctccagtgggcgccgagcccgccctcttcgtgcagggcccccgcgctggctccttc 291  
Db 181 GACACGCTCGCTCGACCGACGCTCCGCGAGGAGATCCGCGCCGACCTGCTCGGCTTC 240  
QY 292 ctgacgcagcagcaagtggcgctgctgctgctgctgctgctgctgctgctgctg 351  
Db 241 GCGCGCGGAGCGGACCACTACGCGCGCTGCGCGGAGATGGAAGCGCGGTCCTC 300  
QY 352 gacagctggagccgctgctgctgctgctgctgctgctgctgctgctgctgctg 388  
Db 301 CACGGCCCGCGGCGGCGGACAGACCACTGCTGTCG 337

RESULT 14  
AG058290 1434 bp DNA linear GSS 02-NOV-2001  
DEFINITION Pan troglodytes DNA, clone: PTB-045C04.R, genomic survey sequence.  
ACCESSION AG058290  
VERSION AG058290.1 GI:16595751  
KEYWORDS GSS; GSS (genome survey sequence).  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male BAC library clone:PTB-045C04.R.  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
REFERENCE 1 (sites)  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE BAC end sequences of Library PTB  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 1434)  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:chimpbes@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS  
Sequencing: M13Rev  
LIBRARY Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI.  
Location/Qualifiers  
1..1434  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-045C04.R"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
BASE COUNT 332 a 509 c 521 g 14 t 58 others  
ORIGIN

Query Match 12.4%; Score 50.6; DB 12; Length 1434;  
Best Local Similarity 48.2%; Pred. No. 34;  
Matches 151; Conservative 0; Mismatches 161; Indels 1; Gaps 1;  
QY 93 ccacctccccgagagcgctgtgacctgggcttgcggcgaggagctgctgcgacgg 152  
Db 244 CCGCCCG 303  
QY 153 gataaacccgagcgagcgagcgacgtccacatcgccccgacgaccccgaggcct 212  
Db 304 GCG 363  
QY 213 gtcggagctgtccatccatcgctccagtggtggcgagcgccctcttccgtgcagcgc 272  
Db 364 GCG 423  
QY 273 ccgcgcgctgtgccttctcgcacgcagcagcaagtcggtgcgctggtcaggac 331  
Db 424 CCG 483  
QY 332 agactctgggtgacttcgagcagcagcctggagcgcgctgcgcaagatcctgcgcgag 391  
Db 484 CG 543  
QY 392 agcagaacgcgcg 404  
Db 544 GCGCGCGCGCGCG 556

RESULT 15  
CNS006XK 935 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
DEFINITION BACR14N09 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL066051  
VERSION AL066051.1 GI:4945019  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 935)  
AUTHORS Direct Submission  
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr) - Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

Source

1. 935  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR14N09"  
/note="end : 17"

BASE COUNT 257 a 170 c 162 g 96 t 250 others  
ORIGIN

Query Match 12.3%; Score 50.2; DB 12; Length 935;  
Best Local Similarity 27.4%; Pred No. 40;  
Matches 94; Conservative 89; Mismatches 160; Indels 0; Gaps 0;

QY 60 gaccgggatacctacggtgctgacattccacattcccgagagcgccctgtgac 119  
|:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|  
Db 860 GSSCCSGCG 801  
QY 120 ctgggcttgcgggagactgctctgacgaggtatcaaccgccgagcgagcgga 179  
|:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|  
Db 800 CGGCG 741  
QY 180 cgtccacatcgcccgaccgagcgagggcctgtcgagcgtctccatccggtccagt 239  
|:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|  
Db 740 SGGSGCSCG 681  
QY 240 gggcgggaccgcccctcttcgtcagggcgcccgccgctggtcgtctctctgacgg 299  
|:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|  
Db 680 CGCCG 621  
QY 300 cagcgacaagtgcgtgcgctcgttcaggagacagactctggtgactctgagagagcgt 359  
|:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|  
Db 620 CCGCSCCG 561  
QY 360 ggagggcgctcggaagatctctcccgagagagagagagagcgc 402  
|:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|  
Db 560 AGSGRMGAGSGRCGGACSGSGGSSSSSKRKGAGAGCGCASCSS 518

Search completed: July 18, 2002, 10:12:56  
Job time: 9976 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: July 18, 2002, 14:15:03 ; Search time 130.38 Seconds  
(without alignments)  
115.010 Million cell updates/sec

Title: US-09-749-185-9  
Perfect score: 692  
Sequence: 1 MSFLVSEELSPKIPVELRYE.....FEDSLEAALGKILAEQNA3 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	692	100.0	135	21	AA1980.DAT
2	606	87.6	135	21	AA1981.DAT
3	544	78.6	135	21	AA1982.DAT
4	446	64.5	135	21	AA1983.DAT
5	80.5	11.6	400	22	AA1984.DAT
6	77	11.1	323	22	AA1985.DAT
7	74	10.7	456	21	AA1986.DAT
8	73	10.7	456	21	AA1987.DAT
9	73	10.5	806	21	AA1988.DAT
10	72.5	10.5	776	22	AA1989.DAT
11	72.5	10.5	777	22	AA1990.DAT

12	72.5	10.5	1369	20	AA1991.DAT
13	72.5	10.5	1369	22	AA1992.DAT
14	71.5	10.3	147	22	AA1993.DAT
15	71	10.3	129	22	AA1994.DAT
16	71	10.3	129	22	AA1995.DAT
17	71	10.3	304	22	AA1996.DAT
18	70.5	10.2	259	22	AA1997.DAT
19	70	10.1	374	22	AA1998.DAT
20	69.5	10.0	146	18	AA1999.DAT
21	69.5	10.0	147	16	AA2000.DAT
22	69.5	10.0	147	21	AA2001.DAT
23	69.5	10.0	494	18	AA2002.DAT
24	69.5	10.0	494	19	AA2003.DAT
25	69.5	10.0	494	19	AA2004.DAT
26	69.5	10.0	494	20	AA2005.DAT
27	69.5	10.0	494	21	AA2006.DAT
28	69	10.0	109	22	AA2007.DAT
29	69	10.0	150	15	AA2008.DAT
30	69	10.0	573	21	AA2009.DAT
31	69	10.0	620	19	AA2010.DAT
32	69	10.0	620	19	AA2011.DAT
33	69	10.0	663	21	AA2012.DAT
34	69	10.0	689	19	AA2013.DAT
35	69	10.0	717	21	AA2014.DAT
36	68.5	9.9	143	21	AA2015.DAT
37	68.5	9.9	338	21	AA2016.DAT
38	68.5	9.9	901	21	AA2017.DAT
39	68.5	9.9	971	22	AA2018.DAT
40	68.5	9.9	971	22	AA2019.DAT
41	68.5	9.9	1043	22	AA2020.DAT
42	68.5	9.9	2626	21	AA2021.DAT
43	68.5	9.9	4643	22	AA2022.DAT
44	68	9.8	309	22	AA2023.DAT
45	67.5	9.8	338	21	AA2024.DAT

ALIGNMENTS

RESULT 1

AA1980.DAT

ID AA1980.DAT standard; Protein; 135 AA.

AC AA1980.DAT

DT 18-APR-2000 (first entry)

DE Streptomyces netropsis SsgA protein.

KW SsgA; liquid culture; filamentous bacteria; secondary metabolite;

KW mycelium; antibiotic; antitumor agent; immunosuppressive agent;

KW hypocholeraemia agent; enzyme inhibitor; antimigraine agent;

KW herbicide; antiparasitic agent; ruminant growth promoter;

KW bioinsecticide; receptor agonist; antagonist; biomass.

OS Streptomyces netropsis.

PN WO200000613-A1.

PD 06-JAN-2000.

PF 25-JUN-1999; 99WO-NL00395.

PR 26-JUN-1998; 98EP-0202148.

PA (UYLE-) RIJKSUNIV LEIDEN.

PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZ.

PI Van Wezel GP, Kraal B, Luiten RGM;

DR WPI; 2000-147269/13.

DR N-PSDB; AA249731.

PT Reducing branching and enhancing fragmentation in filamentous  
PT microorganisms used to improve their liquid culturing properties -  
XX Disclosure; Fig 5; 60pp; English.  
XX

CC The present sequence is *S. netropsis* SsgA protein. SsgA reduces  
CC branching and fragment septation and enhances fragmentation of mycelium  
CC in liquid culture resulting in lower viscosity of culture broths.  
CC Filamentous bacteria can be transformed with ssgA gene-containing  
CC plasmid to enhance the production of secondary metabolites such as,  
CC antibiotics, antitumour agents, immunosuppressive agents,  
CC hypocholesterolaemic agents, enzyme inhibitors, antimigraine agents,  
CC herbicides, antiparasitic agents, ruminant growth promoters,  
CC bioinsecticides, receptor agonists and antagonists and biomass.  
XX  
SQ Sequence 135 AA;

Query Match 100.0%; Score 692; DB 21; Length 135;  
Best Local Similarity 100.0%; Pred. No. 1.5e-72;  
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFLVSEELSKIPVELRYETROPYAVRMFTHLPDGPVTFWAGRELLLDGINRPSGDG 60  
|||||  
DB 1 msflvseelsfripvelryevgdpvairmtfhipgdapvtwafgrelldglnpsgdg 60  
|||||

QY 61 VHTAPTDEGLSDVSIRLQVGADRALFRAGAPPLVAFLDRTDKSVPLGQBOTLGDFEDSL 120  
|||||  
DB 61 vhtaptdeglsvsirlqvgadralfragapplvafldrtksvplgqbotlgdfeds 120  
|||||

QY 121 EAALGKILAEENAG 135  
|||||  
DB 121 eaalgkilaeeqnaq 135  
|||||

RESULT 2  
AA44649  
ID AAY44649 standard; Protein; 135 AA.  
XX  
AC AAY44649;  
XX  
DT 18-APR-2000 (first entry)  
XX  
DE Streptomyces griseus SsgA protein.  
XX  
KW SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.  
XX  
OS Streptomyces griseus.  
XX  
PN W0200000613-A1.  
XX  
PD 06-JAN-2000.  
XX  
PF 25-JUN-1999; 99WO-NL00395.  
XX  
PR 26-JUN-1998; 98EP-0202148.  
XX  
PA (UYLE-) RIJKSUNIV LEIDEN.  
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.  
XX  
PI Van Wezel GP, Kraal B, Luiten RGM;  
XX  
DR WPI; 2000-147269/13.  
DR N-PSDB; AA249727, AA249728.  
XX  
XX Reducing branching and enhancing fragmentation in filamentous  
PT microorganisms used to improve their liquid culturing properties -  
XX Disclosure; Fig 5; 60pp; English.

XX The present sequence is *S. griseus* SsgA protein. SsgA reduces branching  
CC and fragment septation and enhances fragmentation of mycelium in liquid  
CC culture resulting in lower viscosity of culture broths. Filamentous  
CC bacteria can be transformed with ssgA gene-containing plasmid to enhance  
CC the production of secondary metabolites such as, antibiotics, antitumour  
CC agents, immunosuppressive agents, hypocholesterolaemic agents, enzyme  
CC inhibitors, antimigraine agents, herbicides, antiparasitic agents,  
CC ruminant growth promoters, bioinsecticides, receptor agonists and  
CC antagonists and biomass.  
XX  
SQ Sequence 135 AA;

Query Match 87.6%; Score 606; DB 21; Length 135;  
Best Local Similarity 85.9%; Pred. No. 1.5e-62;  
Matches 116; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 MSFLVSEELSKIPVELRYETROPYAVRMFTHLPDGPVTFWAGRELLLDGINRPSGDG 60  
|||||  
DB 1 msflvseelsfripvelryevgdpvairmtfhipgdapvtwafgrelldglnpsgdg 60  
|||||

QY 61 VHTAPTDEGLSDVSIRLQVGADRALFRAGAPPLVAFLDRTDKSVPLGQBOTLGDFEDSL 120  
|||||  
DB 61 vhtaptdeglsvsirlqvgadralfragapplvafldrtksvplgqbotlgdfed 120  
|||||

QY 121 EAALGKILAEENAG 135  
|||||  
DB 121 edalgrilaeeqnaq 135  
|||||

RESULT 3  
AA44651  
ID AAY44651 standard; Protein; 135 AA.  
XX  
AC AAY44651;  
XX  
DT 18-APR-2000 (first entry)  
XX  
DE Streptomyces goldeniensis SsgA protein.  
XX  
KW SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.  
XX  
OS Streptomyces goldeniensis.  
XX  
PN W0200000613-A1.  
XX  
PD 06-JAN-2000.  
XX  
PF 25-JUN-1999; 99WO-NL00395.  
XX  
PR 26-JUN-1998; 98EP-0202148.  
XX  
PA (UYLE-) RIJKSUNIV LEIDEN.  
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.  
XX  
PI Van Wezel GP, Kraal B, Luiten RGM;  
XX  
DR WPI; 2000-147269/13.  
DR N-PSDB; AA249730.  
XX  
XX Reducing branching and enhancing fragmentation in filamentous  
PT microorganisms used to improve their liquid culturing properties -  
XX Disclosure; Fig 5; 60pp; English.  
XX  
CC The present sequence is *S. goldeniensis* SsgA protein. SsgA reduces  
CC branching and fragment septation and enhances fragmentation of mycelium  
CC in liquid culture resulting in lower viscosity of culture broths.

```
CC Filamentous bacteria can be transformed with ssgA gene-containing  
CC plasmid to enhance the production of secondary metabolites such as,  
CC antibiotics, antitumour agents, immunosuppressive agents,  
CC hypocholesterolaemic agents, enzyme inhibitors, antimigraine agents,  
CC herbicides, antiparasitic agents, ruminant growth promoters,  
CC bioinsecticides, receptor agonists and antagonists and biomass.  
XX Sequence 135 AA;  
  
SQ Query Match 78.6%; Score 544; DB 21; Length 135;  
Best Local Similarity 77.8%; Pred.No. 2.5e-55;  
Matches 105; Conservative 12; Mismatches 18; Indels 0; Gaps  
  
Qy 1 MSFLVSELSFKIPVELRYETDPYAVRMTFHLPGDAPVTWAFGRELLLDGINPFGDG 60  
Db | ||||| :||||| :||| :||||||| ||||| :|| ||||  
    1 mslfvselsfrlpvelryetcdpyavrltflpgdapvtwafgrellldgprpcgdgd 60  
  
Qy 61 VHTAPTDEGLSDVSTRIQVGADRALFRAGAPPLVAFLDRTKSVPLGEOITLGDFEDSL 120  
Db ||||| ||| :||||| :||| :||||||| ||||| :||| :  
    61 vhiapadpetfgvirlrvqvgdqamfrvygtaplvaflrdtkivlgdersladfdall 120  
  
Qy 121 EAALGKILAEQNAG 135  
Db | : || : ||||| |||||  
    121 dealdrilaeeqnag 135  
  
RESULT 4  
AAAY44650  
ID AAY44650 standard; Protein; 135 AA.  
AC AAY44650;  
XX AC  
XX DT  
XX DE  
XX STreptomyces albus G SsgA protein.  
XX  
SsgA; liquid culture; filamentous bacteria; secondary metabolite;  
KW mycelium; antibiotic, antitumour agent; immunosuppressive agent;  
KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;  
KW herbicide; antiparasitic agent; ruminant growth promoter;  
KW bioinsecticide; receptor agonist; antagonist; biomass.  
XX OS  
Streptomyces albus G.  
XX OS  
WO200000613-A1.  
XX PN  
XX PD 06-JAN-2000.  
XX PF 25-JUN-1999; 99WO-NL00395.  
XX PR 26-JUN-1998; 98EP-0202148.  
XX  
(UYLE-) RIJKSUNIV LEIDEN.  
PA (NEWE-) NEDERLANDSE ORG WETENSCHAPPELIJKE ONDERZOEKING  
XX (NEW-) NEDERLANDSE ORG WETENSCHAPPELIJKE ONDERZOEKING  
XX Van Wezel GP, Kraal B, Luiten RGM;  
XX  
WPI: 2000-147269/13.  
DR N-PSDB; AAZ49729.  
XX  
XX Reducing branching and enhancing fragmentation in filamentous  
PT microorganisms used to improve their liquid culturing properties -  
XX Disclosure; Fig 5; 60pp; English.  
XX  
The present sequence is S. albus G SsgA protein. SsgA reduces branching  
CC and fragment separation and enhances fragmentation of mycelium in liquid  
CC culture resulting in lower viscosity of culture broths. Filamentous  
CC bacteria can be transformed with ssgA gene-containing plasmid to enhance  
CC the production of secondary metabolites such as, antibiotics, antitumour  
CC agents, immunosuppressive agents, hypocholesterolaemic agents, enzyme  
CC inhibitors, antimigraine agents, herbicides, antiparasitic agents,
```



PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-01322048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132456.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142300.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.

[illegible]

```
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 10.7%; Score 74; DB 21; Length 456;
Best Local Similarity 25.7%; Pred. No. 6.8;
Matches: 26; Conservative 16; Mismatches 35; Indels 24; Gaps 3;

QY 4 LVSELSFKIPVELRYETRDYAVRMTHPLGDPVTVAFGR-----ELLDGINRPSG 57
Db 137 likellsrgvpvdses-----gtplwaaghdqknavevlehnnpna 182
QY 58.DGDVHIAP----TDPEGLSDVSIRLOVGADRALFRAGAPPL 94
Db 183,etednitpllsavaagslscllellvkagakanvfaggatpl 223

RESULT 9
AAV77574
ID AAV77574 standard; Protein; 806 AA.
XX AAV77574;
XX
XX
DT 08-MAY-2000 (first entry)
XX
DE Human cytoskeletal protein (HCVT) (clone 3768043).
XX
KW Human cytoskeletal protein; HCVT; cell proliferation; immunological;
KW reproductive; nervous disorder; cytostatic; immunosuppressive; anti-HIV;
KW anti-diabetic; antiarteriosclerotic; dermatological; anti-inflammatory;
XX anti-infertility; vasotropic; cardiant.
OS Homo sapiens.
XX
PN WO200006730-A2.
PD 10-FEB-2000.
XX
PF 30-JUL-1999; 99WO-US17167.
XX
PR 31-JUL-1998; 98US-0155185.
PR 04-AUG-1998; 98US-0160081.
PR 19-AUG-1998; 98US-0155228.
XX
PA (INCY-) INCYTE PHARM INC.
XX
XX Bandman O, Tang YT, Yue H, Corley NC, Guegler KJ, Azimzai Y;
PI Patterson C, Lal P, Baughn MR;
XX
DR WPI: 2000-195297/17.
XX
PT N-PSDB: AAZ58979.
XX
Human cytoskeletal proteins useful for diagnosing, treating preventing
cell proliferation, immunological, reproductive, developmental and
nervous disorders
```

```

XX PS Claim 1; Page 69-70; 84pp; English.
XX CC
XX CC The invention provides human cytoskeletal proteins (HCYT) and nucleic
XX CC acids encoding the proteins. The HCYT polypeptides can be expressed by
XX CC standard recombinant methodology. The HCYT polypeptides, modulators and
XX CC antibodies are useful for treating or preventing a disorder associated
XX CC with decreased and increased expression or activity of HCYT in mammals.
XX CC The polypeptides are also useful for diagnosing HCYT activity disorders
XX CC such as cell proliferative, immunological, reproductive, developmental
XX CC and nervous disorders. Sequences AAY7569-576 represent HCYT
XX CC polypeptides.
XX SQ Sequence 806 AA;

Query Match 10.5%; Score 73; DB 21; Length 806;
Best Local Similarity 25.0%; Pred. No. 19;
Matches 47; Conservative 13; Mismatches 36; Indels 92; Gaps 10;

Qy 16 ELRYE--TRDPYAVRMTFHLPGDAPVTWAFGRE-----LILLD 50
Db 494 elkfellekdpal-----dvnt-afgrhspypgslpwpssetrflspptlle 54

Qy 51 GINR-----PSGDG-----DVHIAPTDPEGLS---DVS 75
Db 545 gplrlspilpggggrgsgpnpdihqitnergesscdrltdphrapstgslppwdqd 604

Qy 76 IRL-----QVGADRAL-----FRAGAPPLVAF-----LDRTKSVPLGQEQRT 112
Db 605 rirmfpppgsgypdsalppqqrqdrfcnsgrlsgpaelrslrnmpldkmdgmpsemess 664

Qy 113 LGDFEDSL 120
Db 665 rndtkddl 672

RESULT 10
ABG05279
ID ABG05279 standard; Protein: 776 AA.
XX AC ABG05279;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #5270.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS69466.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity
XX DR
```

```

PS Claim 20; SEQ ID No 35638; 103pp; English.
XX CC
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful for medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 776 AA;

Query Match 10.5%; Score 72.5; DB 22; Length 776;
Best Local Similarity 29.0%; Pred. No. 21;
Matches 42; Conservative 13; Mismatches 41; Indels 49; Gaps 9;

Qy 16 ELRYE--TRDPYAVRMTFHLPGDAPVTWAFGRELLLDG-----INRPSGDG-----DV 61
Db 507 elkfellekdpal-----dvnt-afgrgsgpnpdihqitnergesscdrltdp 557

Qy 62 HIAPTDPEGLS---DVSIRL-----QVGADRAL-----FRAGAPPLVAF-- 97
Db 558 hrapsdtgslppwdqdrirmfpppgsgypdsalppqqrqdrfcnsgrlsgpaelrslrnm 617

Qy 98 --LDRTKSVPLGQEQRTLGDFEDSL 120
Db 618 psldkmdgmpsemessrndtkddl 642

RESULT 11
ABG20257
ID ABG20257 standard; Protein: 777 AA.
XX AC ABG20257;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #20248.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR
```



DR N-PSDB; AAS64444.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX

XX

PS ClaI 20; SEQ ID NO 50616; 103pp; English.

PS

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

CC

XX Sequence 777 AA:

XX

```

XX 18-DEC-1998; 98WO-US27140.
XX PF
XX XX
XX 16-DEC-1998; 98US-0212843.
XX PR
XX 20-DEC-1997; 97US-0068379.
XX BR
XX (GEMY ) GENETICS INST INC.
XX PA
XX XX
XX Agostino MJ, Clark HF, Collins-Racie LA, Evans C;
PI Fechtel K, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Steininger RJ, Treacy M, Wong GG;
XX XX
XX WPI; 1999-395405/33.
XX DR
XX N-PSDB; AAX80481.
XX XX
XX New polynucleotides encoding secreted human proteins potentially
PT useful as, e.g. immunostimulators
PT XX
XX Claim 9; Page 84-89; 99pp; English.
XX PS
XX XX
XX The present invention describes human secreted proteins obtained from
CC human fetal brain, fetal kidney or adult blood cDNA libraries. The
CC present sequence represents a human secreted protein. The human secreted
CC proteins, and polynucleotides encoding them, are predicted to have
CC biological activities which would make them suitable for treating,
CC preventing or ameliorating medical conditions in humans and animals,
CC although no supporting data is given. Suggested activities include
CC nutritional activity, cytokine and cell proliferation/differentiation
CC activity, immune stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity, activin/
CC inhibin activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. The polynucleotides are also stated to be useful
CC for gene therapy.
XX XX
XX Sequence 1369 AA;
XX SQ
Query Match 10.5%; Score 72.5; DB 20; Length 1369;
Best Local Similarity 29.0%; Pred. No. 44;
Matches 42; Conservative 13; Mismatches 41; Indels 49; Gaps
QY 16 ELRYE--TRDPYAVRMTFLHPGDPAVTFWAFGRELLLDG-----INRPSGDG-----DV 61
Db 1100 elkfelldkpyal-----dvpt-afgrsgpgnpldhqitnergesscdrltdp 1150
QY 62 HIATDPNGLS---DVSIRL-----QVGADRAL-----FRAGAPPLVAF-- 97
Db 1151 hrapsdtgslspwvqdrrmmfppqgyspdsalppqrqdfcnsgrispaelrsfnn 1210
QY 98 -LDRTKSVPLGQQTGLGDFEDSL 120
Db 1211 psldkmdgsmmpsemessrndtkddl 1235
RESULT 13
AAU39009
ID AAU39009 standard; Protein; 1369 AA.
XX AC
XX AAU39009;
XX XX
XX 16-JAN-2002 (first entry)
XX DE
XX Human secreted protein am728_60.
XX XX
XX Human; secreted protein; antiinflammatory; immunosuppressive;
KW neutropic; neuroprotective; antiarthritic; antimicrobial; vulnery;
KW cystostatic; antidiabetic; virucide; antinfertility; anticonvulsant;
KW vasotrophic; antiparkinsonian; immunostimulant; dermatological;
KW antirheumatic; antitumor; antileucor; osteopathic; tranquiliser;
KW cerebroprotective; cytokine; cell proliferation; cell differentiation;
KW immune deficiency; severe combined immunodeficiency; SCID;tumour;

```

KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;  
KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; infection; cardiac disease;  
KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;  
KW food supplement; vaccine.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175068-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 22-MAR-2001; 2001WO-US09369.  
XX  
XX  
XX 30-MAR-2000; 2000US-0539330.  
PR  
XX  
XX 04-DEC-2000; 2000US-0729674.  
PR  
XX  
XX (GEMY ) GENETICS INST INC.  
PA  
XX  
XX Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;  
PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;  
PI Clark H, Rechtel K, Merberg D;  
XX  
XX WPI; 2001-639363/73.  
DR  
XX  
XX N-PSDB; AA859227.  
XX  
XX Secreted human proteins, useful as vaccine for treating various  
PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and  
PT nervous system disorders (e.g. stroke) -  
PT  
XX  
XX Disclosure; Page 482-487; 619pp; English.  
PS  
XX  
XX The invention relates to novel human secreted proteins, the nucleic  
CC acids encoding them. The protein may exhibit cytokine, cell proliferation  
CC or cell differentiation activity or may induce production of other  
CC cytokines in certain cell populations and may exhibit immune stimulating  
CC or immune suppressing activity, which is useful for the treatment of  
CC various immune deficiencies and disorders e.g. severe combined  
CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,  
CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
CC inflammation. The proteins are also useful in the treatment of diseases  
CC and disorders including tissue, skin and organ transplantation and in  
CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,  
CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,  
CC in the treatment of burns, incisions and ulcers; as well as in treatment  
CC of periodontal disease, osteoporosis or osteoarthritis, mediated by  
CC inflammatory processes, diseases of the peripheral nervous system,  
CC Alzheimer's, Parkinson's disease, Huntington's disease,  
CC amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections,  
CC infarction of cardiac and central nervous system vessel e.g. stroke,  
CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The  
CC protein, having activin- or inhibin-related activities is useful as a  
CC contraceptive based on the ability of inhibins to decrease fertility in  
CC female mammals and decrease spermatogenesis in male mammals. The  
CC proteins and nucleic acids are also useful as food supplements. The  
CC present sequence represents a secreted protein of the invention.  
XX  
XX  
XX Sequence 1369 AA;  
SQ

Query Match 10.5%; Score 72.5; DB 22; Length 1369;  
Best Local Similarity 29.0%; Pred. No. 44;  
Matches 42; Conservative 13; Mismatches 41; Indels 49; Gaps 9;  
QY 16 ELRYE--TRDPYAVRWTFHLPDAPVTWAFGRLLLDG-----INRPSGDG-----DV 61  
Db 1100 elkfellekdpal-----dvpnt-afgrsgpgnpldqhtnrgesscdrltdp 1150  
QY 62 HIAPTDPEGLS---DVSIRL-----QVGADRAL-----FRAGAPPLVAF-- 97  
Db 1151 hrapsdtgslppwdqrrmmfpppgsydpdsalppqrqdrfcnsgrlsgpaelrsfnn 1210

QY 98 --LDRTDKSVPLGQEQTLGDFEDSL 120  
Db 1211 psldkmdgmpseessrntdkdl 1235  
RESULT 14  
AAB98959  
ID AAB98959 standard; Protein; 147 AA.  
XX  
XX AAB98959;  
AC  
XX 20-AUG-2001 (first entry)  
DT  
XX Murine Mphi type Iba1.  
DE  
XX Human; mouse; immunomodulatory; monocyte; macrophage; inhibitor.  
KW  
XX Mus sp.  
OS  
XX JP2001078775-A.  
PN  
XX 27-MAR-2001.  
PD  
XX 14-SEP-1999; 99JP-0260793.  
PF  
XX 14-SEP-1999; 99JP-0260793.  
PR  
XX (KOKU-) KOKURITSU SEISHIN SHINKEI CENT SOCHO.  
PA (YAK-) IYAKUHN KUKUSAYO HIGAI KYUSAI KENKYU SH.  
PA (MOCH) MOCHIDA PHARM CO LTD.  
XX  
XX WPI; 2001-313369/33.  
DR  
XX N-PSDB; AAH25799.  
XX  
XX A macrophage function modifier useful for preventing and treating  
PT diseases caused by the increase or decrease in macrophage activity or  
PT function -  
XX  
XX Claim 10; Page 12; 20pp; Japanese.  
PS  
XX The present invention provides a number of murine and human Iba1  
CC derivatives, which are capable of inhibiting the function of cells with  
CC monocyte or macrophage activity. These can be used as immunomodulators  
CC to prevent and treat diseases caused by a decrease or increase in the  
CC activity or the function of macrophages or an activator or an inhibitor  
CC of the function of cells of macrophage type. The present sequence is one  
CC of the derivatives of the invention.  
XX  
XX Sequence 147 AA;  
SQ

Query Match 10.3%; Score 71.5; DB 22; Length 147;  
Best Local Similarity 28.1%; Pred. No. 2; 9;  
Matches 34; Conservative 12; Mismatches 42; Indels 33; Gaps 6;  
QY 42 AFG-----RELLDGINRP-----SGDGD-----VHTAPDPEGLSDVSRILQV 80  
Db 12 afgllkaqqeergleginkqflddpkysndedlpskleafkvkymefdingngdidi---m 68  
QY 81 GADRALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDSLAAALGK-----ILAEQ 132  
Db 69 slkrmiekgvpxthlelrlrevssgseetf-sysdfmrmlgkrsallrmlmyeek 127  
QY 133 N 133  
Db 128 n 128  
RESULT 15  
AAG90040  
ID AAG90040 standard; Protein; 129 AA.  
XX  
XX AAG90040;  
AC

Search completed: July 18, 2002, 14:15:04  
Job time: 10334 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:16:08 ; Search time 51.31 Seconds  
(without alignments)  
64.265 Million cell updates/sec

Title: US-09-749-185-9

Perfect score: 692

Sequence: 1 MSFLVSELSFKIPVELRYE.....FEDSLEALGKILAEQNAQ 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	79.5	11.5	3072	4	US-09-413-814-93
2	79.5	11.5	3079	4	US-09-413-814-80
3	75	10.8	881	4	US-09-413-814-44
4	70.5	10.2	259	1	US-08-015-986A-10
5	70.5	10.2	259	1	US-08-015-973-6
6	70.5	10.2	259	2	US-08-446-363-10
7	70.5	10.2	259	2	US-08-448-164-6
8	70.5	10.2	259	4	US-08-081-929-6
9	69.5	10.0	146	4	US-08-792-013-7
10	69.5	10.0	147	1	US-08-171-385-5
11	69.5	10.0	147	3	US-08-361-441B-5
12	69.5	10.0	474	1	US-08-485-618-103
13	69.5	10.0	494	2	US-08-605-672-103
14	69.5	10.0	494	2	US-08-482-293A-103
15	69.5	10.0	494	2	US-08-943-363-103
16	69.5	10.0	494	4	US-09-193-043-103
17	68.5	9.9	143	3	US-08-685-808-2
18	68.5	9.9	143	4	US-08-505-860C-2
19	68	9.8	527	4	US-08-984-618-9
20	66.5	9.6	576	4	US-09-367-206-1
21	66.5	9.6	576	4	US-09-367-206-21
22	66.5	9.6	576	4	US-09-367-206-22
23	66.5	9.6	576	4	US-09-367-206-23
24	66	9.5	150	1	US-08-015-770B-8
25	64.5	9.3	942	3	US-09-074-579-1
26	64.5	9.3	942	4	US-09-388-774-1
27	64	9.2	987	1	US-08-436-044-6

28	64	9.2	987	2	US-08-436-054-6
29	64	9.2	987	5	PCT-US95-08812-6
30	64	9.2	1276	1	US-08-222-616-24
31	64	9.2	1276	5	PCT-US95-04228-24
32	63	9.1	816	2	US-08-267-803B-9
33	63	9.1	816	4	US-09-041-886-17
34	63	9.1	1090	4	US-09-346-237-5
35	62.5	9.0	311	4	US-09-056-556-188
36	62.5	9.0	330	3	US-08-851-843A-203
37	62.5	9.0	330	4	US-08-974-549A-322
38	62.5	9.0	330	4	US-08-854-050-203
39	62.5	9.0	330	4	US-09-430-323-203
40	62.5	9.0	337	1	US-08-312-387B-5
41	62.5	9.0	337	1	US-08-312-387B-12
42	62.5	9.0	337	1	US-08-683-426-5
43	62.5	9.0	337	1	US-08-683-426-12
44	62.5	9.0	337	1	US-08-683-458-5
45	62.5	9.0	337	1	US-08-683-458-12

ALIGNMENTS

RESULT 1

US-09-413-814-93  
; Sequence 93, Application US/09413814  
; Patent No. 6225064

; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bloecker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hofle, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans

; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
; TITLE OF INVENTION: heteropolyketide compounds  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413.814  
; CURRENT FILING DATE: 1999-10-07  
; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
; EARLIER FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 93  
; LENGTH: 3072  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-413-814-93

Query Match 11.5% Score 79.5; DB 4; Length 3072;  
Best Local Similarity 31.5%; Pred. No. 4.2; Indels 11; Gaps 3;  
Matches 35; Conservative 8; Mismatches 57;

QY 1 MSFLVSELSFKIPVELRYETRDYAVRMTFHLPCDAPVTWAFGRLLDGINRPSGDG 60

Db 869 MOVLGRIESSIGIRTLTLEFHTLAQLASHLSGSAASTSAAATALERGLTRPDGFS 928

QY 61 VHIA-PTDP-----EGLSDVSIQLQVADRALF-----RAGAPPLVAFLD 100

Db 929 PRVATPEEPFALTEGQAMWLECOKSADGALYNLGRTRVRLGAGVDVAALRR 979

RESULT 2

US-09-413-814-80  
; Sequence 80, Application US/09413814  
; Patent No. 6225064

; GENERAL INFORMATION:

```

; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 3079
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-80

Query Match 11.5%; Score 79.5; DB 4; Length 3079;
Best Local Similarity 31.5%; Pred. No. 4.2;
Matches 35; Conservative 8; Mismatches 57; Indels 11; Gaps 3;

QY 1 MSPFLVEELSFKIPVELRYETRPYAVRMTFHLPGDAPVTWAFGRELDDGINRPSGDGD 60
Db 875 MQVLGRIESLGRITRTLRFHEPTLAQLASHSSGAASSTAATALERTGTRDPGSS 934

QY 61 VHIA-PTDP-----EGLSDVSIRLQVGADRALF-----RAGAPPLVAFIDR 100
Db 935 PRVATPEEPFALTEGORAMWLECKSADGALYNLGRTRVLGADVAAALR 985

RESULT 3
US-09-413-814-44
; Sequence 44, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 881
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-44

Query Match 10.8%; Score 75; DB 4; Length 881;
Matches 27; Conservative 18; Mismatches 34; Indels 33; Gaps 6;

Best Local Similarity 29.0%; Pred. No. 2.6;
Matches 38; Conservative 13; Mismatches 44; Indels 36; Gaps 9;

QY 15 VELRYETRPYAVRMTFHLPGDAPVTWAFGRELDDGINRPSGDGDVHIAPTD 67
Db 594 LDRGRRQPPY-----LPG-----ELFLAGDCLARGYLNRPDITA-LHFVP-N 634

QY 68 PEGLSDVSIQLQVGADRALFRAGAPPLVAFIDRTDKSVPL-GQEQTLDGFEDSLEA---- 122
Db 635 PFGNGE---RMYHSGDLALVRGDGQ--VAFGLRRDHIQIRGORVELGETESHRLGLEGI 689

QY 123 ALGKILAEQN 133
Db 690 AAADVQAESQH 700

RESULT 4
US-08-015-986A-10
; Sequence 10, Application US/08015986A
; Patent No. 5532123
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-GAMMA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/015,986A
; FILING DATE: 10-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 259 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-015-986A-10

Query Match 10.2%; Score 70.5; DB 1; Length 259;
Best Local Similarity 24.1%; Pred. No. 1.7;
Matches 27; Conservative 18; Mismatches 34; Indels 33; Gaps 6;

QY 24 PYAVRMTFHLPGDAPVTWA----FGRELLDGINRPSGDGDVHIAPTD----- 68
Db 86 PYRLRQ-PHL-----HWGSSDDHGHSEHTVDGVKYAA----ELHLVHWNPKYNTFKREALQ 135

QY 69 -EGLSDVSIRLQVGADRALFRAGAPPLVAFIDRTDKSVPLGQEQTLGDFEDS 119
Db 136 RDSIAVIGIFLKGHENGFEQ-----IFLDALDKIKTKGKEAPPTKFDPS 180

RESULT 5
```



US-08-081-929-6

US-08-081-929-6

```

Query Match      10.0%; Score 69.5; DB 4; Length 146;
Best Local Similarity 27.3%; Pred. No. 1;
Matches 33; Conservative 13; Mismatches 42; Indels 33; Gaps 6;

QY 42 AFG-----RELLDGINRP-----SGDGDVH-----IAPTDPEGLSDVSIRLQV 80
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 11 AFGLLQAQGEERLDGNKHFLLDPPKYSSDEDLQSLKLEAFKTYMEFDLNGNDIDI---M 67

QY 81 GADRALFRACAPPLVAPLFDTRXSVLPQGEQTLGDFEDSLAALGK-----ILAEEQ 132
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 68 SLKRMLEKGVKPTHLRLKLLIREVSSGSETF--SYSDFLRMMLGKRSAIRLMLMVEEK 126
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

Best Local Similarity 27.3%; Pred No. 1:
Matches 33; Conservative 13; Mismatches 42; Indels 33; Gaps 6;

Qy 42 AFG-----RELLLDGINRP-----SGDGDVH-----IAPDPEGLSDVSIRLQV 80
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 11 AFGLLQAQOEERLDGINKHFLDPPKYSSDEEDLQSKLEAFKTKMYEFDLNGDIDI--M 67

Qy 81 GADRALFRAGAPPLVAFLDRTDTSVPVQGBOTLGDPDFDSLEAALGK-----ILAEQ 132
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 68 SLKRMLEKGVPPKTHLEKLKLIREVSSGSSETF--SYSDFLRMMLGKRSAIRLMLMVEEK 126
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

68 SLKRMLEKLGVPKTHLELKKLIREVSSGSEETF-SYSDFLRMMLGKRSAILRMILMYEEK 126



```

QY 133 N 133
Db 127 N 127

RESULT 10
US-08-171-385-5
; Sequence 5, Application US/08171385
; Patent No. 5527884
; GENERAL INFORMATION:
; APPLICANT: Mary E. Russell
; APPLICANT: Ulrike Utans
; TITLE OF INVENTION: Mediators of Chronic Allograft
; TITLE OF INVENTION: Rejection
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,385
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/006001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 147
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-171-385-5

Query Match 10.0%; Score 69.5; DB 1; Len
Best Local Similarity 27.3%; Pred. No. 1;
Matches 33; Conservative 13; Mismatches 42; I
QY 42 AFG-----RELLDGINRP-----SGDGDVH-----IAPTT
Db 12 AFGLLKAQCEERLDGINKHFLLDDPKYSDEDLQSKLEAFKTKYMEFF

QY 81 GADRALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDSLEAALGK
Db 69 SLKRMEKLGVPKTHLEKLLIREVSSGETF-SYSDFLRMMLGK

QY 133 N 133
Db 128 N 128

RESULT 11
US-08-361-441B-5
; Sequence 5, Application US/08361441B
; Patent No. 6077948

```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-103

Query Match 10.0%; Score 69.5; DB 1; Length 494;  
Best Local Similarity 23.4%; Pred. No. 5.3;  
Matches 33; Conservative 17; Mismatches 54; Indels 37; Gaps 7;  
  
QY 2 SFLVSELSFKIPVELRYETRDYPVAVRMTHFLPGDAPVTWAFGRELDDGINRPSGDGD- 60  
DB 123 TLLVGLSLELTVTVTVRNEDSGSYGTAITYYPAGL-----SYRVSGQTQP 169  
  
QY 61 -----VHIA-----PTDPEGLSDVSIRLQVGADRALFRAGAPPLVAFLDRTDKSVPLGQEQT 112  
DB 170 WORPLHLACEAVPTSEGLRST-----CSVNHPIFOGGAQG--TFVVKFDVS-----SKAS 219  
  
QY 113 LGDFEDSLEALGKILAEQON 133  
DB 220 LGD-----RLMGASASSEN 235

RESULT 13  
US-08-605-672-103  
Sequence 103, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497

APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-103

Query Match 10.0%; Score 69.5; DB 2; Length 494;  
Best Local Similarity 23.4%; Pred. No. 5.3;  
Matches 33; Conservative 17; Mismatches 54; Indels 37; Gaps 7;  
  
QY 2 SFLVSELSFKIPVELRYETRDYPVAVRMTHFLPGDAPVTWAFGRELDDGINRPSGDGD- 60  
DB 123 TLLVGLSLELTVTVTVRNEDSGSYGTAITYYPAGL-----SYRVSGQTQP 169  
  
QY 61 -----VHIA-----PTDPEGLSDVSIRLQVGADRALFRAGAPPLVAFLDRTDKSVPLGQEQT 112  
DB 170 WORPLHLACEAVPTSEGLRST-----CSVNHPIFOGGAQG--TFVVKFDVS-----SKAS 219  
  
QY 113 LGDFEDSLEALGKILAEQON 133  
DB 220 LGD-----RLMGASASSEN 235

RESULT 14  
US-08-482-293A-103  
Sequence 103, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497

; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 103:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 494 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-482-293A-103

Query Match 10.0%; Score 69.5; DB 2; Length 494;  
Best Local Similarity 23.4%; Pred. No. 5.3;  
Matches 33; Conservative 17; Mismatches 54; Indels 37; Gaps 7;  
QY 2 SFLVSEELSKIPVELRYETRDPIYAVMTFHLPGDAPVTWAFGRELLLDGINRPSGDGD- 60  
DB 123 TLLVGLSLTLVTVTRNEGEDSYGTAITLYYPAGL-----SYRRVSGQTQP 169  
QY 61 ----VHIA----PTDPEGLSDYSIRLOVGADRALFRAGAPPLVAFLDRTDKSVPLQEQOT 112  
DB 170 WQPHLACEAVPTSEGLRST-----CSVNHPIFOGGAQG--TFVVKFDVS-----SKAS 219  
QY 113 LGDFEDSLAALGKILAEQN 133  
DB 220 LGD-----RLMGASASSEN 235

RESULT 15  
US-08-943-363-103  
; Sequence 103, Application US/08943363  
; Patent No. 5837478  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/943,363  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 103:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 494 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-943-363-103

Query Match 10.0%; Score 69.5; DB 2; Length 494;  
Best Local Similarity 23.4%; Pred. No. 5.3;  
Matches 33; Conservative 17; Mismatches 54; Indels 37; Gaps 7;  
QY 2 SFLVSEELSKIPVELRYETRDPIYAVMTFHLPGDAPVTWAFGRELLLDGINRPSGDGD- 60  
DB 123 TLLVGLSLTLVTVTRNEGEDSYGTAITLYYPAGL-----SYRRVSGQTQP 169  
QY 61 ----VHIA----PTDPEGLSDYSIRLOVGADRALFRAGAPPLVAFLDRTDKSVPLQEQOT 112  
DB 170 WQPHLACEAVPTSEGLRST-----CSVNHPIFOGGAQG--TFVVKFDVS-----SKAS 219  
QY 113 LGDFEDSLAALGKILAEQN 133  
DB 220 LGD-----RLMGASASSEN 235

Search completed: July 18, 2002, 14:16:09  
Job time: 9699 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:17:37 ; Search time 73.98 seconds  
(without alignments)  
175.345 Million cell updates/sec

Title: US-09-749-185-9

Perfect score: 692

Sequence: 1 MSFLVSEELSFKIPVELRYE.....FEDSLAALGKILAEQNAG 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	548	79.2	136	2	T37179	probable regulator
2	237	34.2	142	2	T36147	probable regulator
3	159	23.0	138	2	T35247	probable regulator
4	132.5	19.1	142	2	T35319	probable regulator
5	80	11.6	514	2	G81935	amidophosphoribosyl transferase
6	79.5	11.5	259	2	A22612	carbonate dehydratase
7	79	11.4	1101	2	S15271	endoglucanase C (E)
8	76.5	11.1	313	2	H84169	hypothetical protein
9	76	11.0	514	2	A81170	amidophosphoribosyl transferase
10	75.5	10.9	818	2	F87327	conserved hypothetical protein
11	75.5	10.9	902	2	G83635	probable ClpA/B-type
12	75	10.8	611	1	S12566	translation initiation factor
13	74.5	10.8	260	2	I52551	carbonate dehydratase
14	74.5	10.8	291	2	H70678	hypothetical protein
15	74.5	10.8	408	2	A33237	conserved hypothetical protein
16	74.5	10.8	757	2	G97472	penicillin-binding protein
17	74.5	10.8	757	2	AC2691	hypothetical protein
18	74	10.7	347	2	B87500	conserved hypothetical protein
19	73.5	10.6	260	2	A43641	carbonate dehydratase
20	73.5	10.6	346	1	H89789	probable alcohol dehydrogenase
21	73.5	10.6	732	2	B90668	hypothetical protein
22	73.5	10.6	732	2	E85518	hypothetical protein
23	73.5	10.6	891	2	A80358	Clp ATPase (import)
24	73	10.5	265	2	C70963	hypothetical protein
25	73	10.5	1337	2	B64993	hypothetical protein
26	73	10.5	1534	2	G91017	probable membrane protein
27	73	10.5	1534	2	A85862	hypothetical protein
28	72.5	10.5	343	2	I69009	MHC class I RT1.C
29	72.5	10.5	657	2	E75484	glycosyl hydrolase

ALIGNMENTS

RESULT 1

T37179

probable regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T37179

R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: Z21598

A:Accession: T37179

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-136 <SEE>

A:Cross-references: EMBL:AL096823; PIDN:CAB46964.1; GSPDB:GN00070; SCOEDB:SCQ11.09

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: ssgA; SCOEDB:SCQ11.09

Query Match 79.2%; Score 548; DB 2; Length 136;

Best Local Similarity 77.8%; Pred. No. 1.le-46;

Matches 105; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MSFLVSEELSFKIPVELRYETRDYAVRTHLPDAPVTWAFGRLLDGNRPSGCDG 60

Db 2 MSFLVSEELSFKIPVELRYETRDYAVRTHLPDAPVTWAFGRLLDGVGRCGDDG 61

Qy 61 VHIAPTPDPEGLSDVSIQLQVGADRALFRAGAPLVAFLDRTKSVPLGQEQTLGDFEDSL 120

Db 62 VRIAPVEPEPLAEVLIRLQVGSQDALFRSSAAPLVAFLDRTKLVPLGQEGALADFDL 121

Qy 121 EAALGKILAEQNAG 135

Db 122 DEALDRILAEQNSAG 136

RESULT 2

T36147

probable regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000

C:Accession: T36147

R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: Z21598

A:Accession: T36147

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-142 <SEE>

A:Cross-references: EMBL:AL096852; PIDN:CAB51005.1; GSPDB:GN00070; SCOEDB:SCE19A.24

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC19A.24  
C:Superfamily: Streptomyces coelicolor probable regulator SCE19A.24

Query Match 34.2%; Score 237; DB 2; Length 142;  
Best Local Similarity 39.7%; Pred. No. 4.1e-16;  
Matches 50; Conservative 20; Mismatches 56; Indels 0; Gaps 0;  
  
QY 1 MSFLVSELSFKIPVELRYETRDYAVRMTFHLPGDAPVTWAFGRELLLDGINRPSGDGD 60  
DB 17 LRLVLPESGTPVPAALGYHTDDPYAVHITHSDGHPVHTFADLLVEGVFRPSGHGD 76  
QY 61 VHIAPTDEGLSDVSIQLQVADRALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDSL 120  
DB 77 VRVPSKTEGRSVVLVALSSPDGDLLEAPTQVSAWLERTLRAVPPGTEGAQLGIDGL 136  
QY 121 EAAALGK 126  
DB 137 AELLAR 142

RESULT 3  
T35247  
Probable regulator - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 20-Jun-2000  
C:Accession: T35247  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, April 1999  
A:Reference number: 221573  
A:Accession: T35247  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-138 <OLI>  
A:Cross-references: EMBL:AL049587; PIDN:CAB40672.1; GSPDB:GN00070; SCOEDB:SC5F2A.05c  
A:Experimental source: strain A3(2)  
C:Genetics:  
C:Superfamily: Streptomyces coelicolor probable regulator SCE19A.24

Query Match 23.08; Score 159; DB 2; Length 138;  
Best Local Similarity 33.1%; Pred. No. 1.8e-08;  
Matches 40; Conservative 13; Mismatches 42; Indels 26; Gaps 3;  
  
QY 13 IPVELRYETRDYAVRMTFHLPG--DAPVTWAFGRELLLDGINRPSGDGDVHIAP---- 65  
DB 23 IPATLHYDRADPFAVRMTFPAPATLEGVEVCWTFESRELLIAGMOEPNGHGVDVRVPYAYD 82  
QY 66 -----TDPEGLSDVSIQLQVADRALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDS 119  
DB 83 RTVLEFHAPEGTAIVHVR-----SGELRFLQAGELVPVGLLEHLQLDLHD 129  
QY 120 L 120  
DB 130 L 130

RESULT 4  
T35319  
Probable regulator - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 20-Jun-2000  
C:Accession: T35319  
R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: 221575  
A:Accession: T35319  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-142 <OLI>  
A:Cross-references: EMBL:AL049863; PIDN:CAB42928.1; GSPDB:GN00070; SCOEDB:SC5H1.03  
A:Experimental source: strain A3(2)

C:Genetics:  
A:Gene: SCOEDB:SC5H1.03  
C:Superfamily: Streptomyces coelicolor probable regulator SCE19A.24

Query Match 19.1%; Score 132.5; DB 2; Length 142;  
Best Local Similarity 30.6%; Pred. No. 7.6e-06;  
Matches 37; Conservative 21; Mismatches 52; Indels 11; Gaps 4;  
  
QY 11 FKIPVELRYETRDYAVRMTFHLPGDAPVTWAFGRELLLDGINRPSGDGDVHIAPTDEG 70  
DB 22 FPLVLAHLAYDAADPFALTVPFVSHDGRVLARLWTLDRVMAEGLTRPVGVGDVRLRP-ESRG 80  
QY 71 LSDVSIQLQVAD-----RALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDSLE 121  
DB 81 MWD-ELRIELLDGRADGERHRAVVFVWAAVAEFLRETHAVVRPGREVRVDDFLAELT 139  
QY 122 A 122  
DB 140 A 140

RESULT 5  
G81935  
amidophosphoribosyltransferase (EC 2.4.2.14) NMA0892 [similarity] - Neisseria meningi  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: G81935  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A:Reference number: A81775; MUID:20222556  
A:Accession: G81935  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-514 <PAR>  
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84172.1; PID:g737  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: purF; NMA0892  
C:Superfamily: amidophosphoribosyltransferase  
C:Keywords: glycosyltransferase; pentosyltransferase  
F:2-514/Product: amidophosphoribosyltransferase #status predicted <MAT>  
F:2/Active site: Cys #status predicted

Query Match 11.6%; Score 80; DB 2; Length 514;  
Best Local Similarity 23.0%; Pred. No. 5.1;  
Matches 50; Conservative 24; Mismatches 53; Indels 90; Gaps 11;  
  
QY 1 MSFLVSELSFKIPV-----ELRYETRDY-----VRMTFHLP 34  
DB 284 MGVSLAEKIKRELDPVDGIDVYMPIDPSTRPSAMELAHLKKVPYREGLIKRYTGRTIMP 343  
QY 35 GDA-----PVTWAF-GRELL--DGI-----NRPSGDGDVHIAPTDP 68  
DB 344 GQATRKSVRQKLSPMETEFAGKSVLLVDDSIYRGTTSRIVEVMVRAAGARKYVIAAAP 403  
QY 69 E-----GLSDVSIQLQVAD-----RALFRAGAPPLVAF 98  
DB 404 EVRYPNVYGDIMPTRELIANGRSAAEIAAEIGADGIVFDGLDGLAVVKALNPKIESF- 462  
QY 99 DRTDKSVPLGQEQTLGDFEDSLEALGKILAEQNAAG 135  
DB 463 ---DSSCFNGIYQT-GDIDD---AYDLRLSAEKSGCG 492

RESULT 6  
A22612  
carbonate dehydratase (EC 4.2.1.1) III - horse  
N:Alternate names: carbonate dehydratase III  
C:Species: Equus caballus (domestic horse)

C>Date: 22-Jul-1987 #sequence\_revision 22-Jul-1987 #text\_change 25-Apr-1997  
C:Accession: A22612  
R:Wendorff, K.M.; Nishita, T.; Jabusch, J.R.; Deutsch, H.F.  
J. Biol. Chem. 260, 6129-6132, 1985  
A:Title: The sequence of equine muscle carbonic anhydrase.  
A:Reference number: A22612; MUID:85207593  
A:Accession: A22612  
A:Molecule type: protein  
A:Residues: 1-259 <WEN>  
C:Superfamily: carbonic dehydratase; carbonic anhydrase homology  
C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc  
F:4-258/Domain: carbonic anhydrase homology <CAH>  
F:1/Modified site: acetylated amino end (Ala) #status predicted

Query Match 11.5%; Score 79.5; DB 2; Length 259;  
Best Local Similarity 26.6%; Pred. No. 2.5; Mismatches 16; Indels 27; Gaps 5;  
Matches 29; Conservative 16

Qy 24 PYAVRMTFPLPDAPVTWA-----FGRELLLDGINRPSGDGVHIAPI-----TDPGE 70  
Db 86 PYRLRQ-PHL--;---HWSSDDHGEHTVDGVKYAAELHLVHNPKYNTYGGALKQPDG 138  
Qy 71 LSDVIRIQVGADRALFRAGAPPIVAFIDRTDKSVPLGQEQTLGDFDS 119  
Db 139 IAVVGVELKIGKEGFEQ-----LFELDALKIKTKGKEAPFTNFDPDS 180

RESULT 7  
S15271  
endoglucanase C (EC 3.2.1.-) - Cellulomonas fimi  
C:Species: Cellulomonas fimi  
C>Date: 21-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 01-Dec-2000  
C:Accession: S15271; A43636  
R:Coutinho, J.B.; Moser, B.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr., R.C.  
Mol. Microbiol. 5, 1221-1233, 1991  
A:Title: Nucleotide sequence of the endoglucanase C gene (cenC) of Cellulomonas fimi, its  
A:Reference number: S15271; MUID:92065819  
A:Accession: S15271  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1101 <COU>  
A:Cross-references: EMBL:X57858; NID:948847; PIDN:CAA040993.1; PID:g580980  
A:Note: The authors translated the codon GAC for residues 361, 380, 400, 550, 670, and 8  
R:Moser, B.; Gilkes, N.R.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr., R.C.  
Appl. Environ. Microbiol. 55, 2480-2487, 1989  
A:Title: Purification and characterization of endoglucanase C of Cellulomonas fimi, clone  
A:Reference number: A43636; MUID:90103465  
A:Accession: A43636  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-64, 'W' <MOS>  
A:Cross-references: GB:M29707; NID:g144417; PIDN:AAA23087.1; PID:g144420  
C:Genetics:  
A:Gene: cenC  
A:Start codon: GTG  
C:Keywords: glycosidase; hydrolase

Query Match 11.4%; Score 79; DB 2; Length 1101;  
Best Local Similarity 30.5%; Pred. No. 16;  
Matches 25; Conservative 12; Mismatches 35; Indels 10; Gaps 4;

Qy 19 YETRPYAV--RMTHFLPGDAPVTWAFGRELLELDGINRPSGDGVHIAPIPTDPEGLSDYSI 76  
Db 200 YGTSEPFVFDAGMCDVLDPGGQGNPDAG--LVYNGV--PVGEGESYVLSFTASATPDMPV 255  
Qy 77 RLQV----GADRALFRAGAPPL 94  
Db 256 RVLVCGEGGAYTAEQGSAPL 277

RESULT 8

H84169  
hypothetical protein Vng0096c [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: H84169  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84169; MUID:20504483  
A:Accession: H84169  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-313 <STO>  
A:Cross-references: GB:AE004437; NID:g10579744; PIDN:AAG18724.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG0096C

Query Match 11.1%; Score 76.5; DB 2; Length 313;  
Best Local Similarity 24.4%; Pred. No. 6.2; Mismatches 19; Indels 53; Gaps 8;  
Matches 41; Conservative 19

Qy 11 FKIPVELRYET-----RDPYAVRMTFHLPGDAPVTWAFGRELLLD--GINRPS 56  
Db 113 FTTGLDADYETIEAHCEADVLAQVADATEVRYTPTLGTD--ITFEAGDRAWDDTGIVRES 170  
Qy 57 GD-----GDVHIAPTDEG-----LSDVIRIQVGADRALFRAGAPPLVAF 97  
Db 171 GFNSLPAGEVFPSTPDANGTYVVDGTMMPHGKLDDELRFEV-ADGVVTEISD----- 223  
Qy 98 LDRDKSVPLGQEQTLGDFEDSLEAL-----GKILAEQNAQ 135  
Db 224 -DRVREQVEAGSEEVGEDAYNLAELGIGNVAVTELGVSLDDEKAAG 270

RESULT 9  
A81170  
amidophosphoribosyltransferase (EC 2.4.2.14) NMB0690 [similarity] - Neisseria meningi  
C:Species: Neisseria meningitidis  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: A81170  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignanl, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755  
A:Accession: A81170  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-514 <TET>  
A:Cross-references: GB:AE002423; GB:AE002098; NID:g7225913; PIDN:AAF41108.1; PID:g722  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0690  
C:Superfamily: amidophosphoribosyltransferase  
C:Keywords: glycosyltransferase; pentosyltransferase  
F:2-514/Product: amidophosphoribosyltransferase #status predicted <MAT>  
F:2/Active site: Cys #status predicted

Query Match 11.0%; Score 76; DB 2; Length 514;  
Best Local Similarity 23.0%; Pred. No. 13;  
Matches 50; Conservative 23; Mismatches 54; Indels 90; Gaps 11;

Qy 1 MSFLVSEELSFKIPV-----ELRYETRDYPA-----VRMTFHLPL 34  
Db 284 MGVSLAEKIKRELKRELPDGVDMVPIPDTSRPSAMELAVHLDKPYREGLIKNRVIGRTFTIMP 343

15



Query Match 10.8%; Score 74.5; DB 2; Length 291;  
Best Local Similarity 25.9%; Pred. No. 8.9;  
Matches 36; Conservative 20; Mismatches 40; Indels 43; Gaps 7;

Search completed: July 18, 2002, 14:17:38  
Job time: 9548 sec

Query Match	10.8%;	Score 74.5;	DB 2;	Length 291;
Best Local Similarity	25.9%;	Pred. No. 8.9;		
Matches 36;	Conservative	20;	Mismatches 40;	Indels 43;
				Gaps 7;

**THIS PAGE BLANK (USPTO)**

Result No.	Query			ID	Description		
	Score	Match	Length				
1	79.5	11.5	259	1	CAH3_HORSE	P07450	equus cabal
2	79	11.4	1101	1	GUNC_CELFI	P14090	cellulomona
3	75	10.8	611	1	IPAB_HUMAN	P23588	homo sapien
4	74.5	10.8	259	1	CAH3_RAT	P14141	rattus norv
5	73.5	10.6	259	1	CAH3_MOUSE	P16015	mus musculu
6	73	10.5	432	1	YQ31_MYCTU	P71930	mycobacteri
7	72.5	10.5	732	1	YAGR_ECOLI	P77489	escherichia
8	71	10.3	554	1	SECD_RHOCA	O33517	rhodobacter
9	70.5	10.2	259	1	CAH3_HUMAN	P07451	homo sapien
10	70.5	10.2	3718	1	LNA5_MOUSE	O61001	mus musculu
11	70	10.1	690	1	VATI_DEIRA	Q9FW33	deinococcus
12	69.5	10.0	147	1	A1F1_RAT	P95009	rattus norv
13	69.5	10.0	338	1	CYF_PHOLA	P95522	phormidium
14	69.5	10.0	484	1	HKK1_SCHPO	O09756	schizosacch
15	69.5	10.0	602	1	Y717_MYCTU	O10966	mycobacteri
16	69	10.0	620	1	EXON_HSV2	P05489	herpes simp
17	68.5	9.9	339	1	ENDA_HALVO	O07118	halobacteri
18	68.5	9.9	469	1	GLN1_STRVR	O05542	streptomyce
19	68	9.9	527	1	RF3_HAEIN	P43928	haemophilus
20	67.5	9.8	1114	1	E2K3_MOUSE	Q92zb5	mus musculu
21	67	9.7	252	1	PS73_DROME	O27575	dirosophila
22	66.5	9.6	143	1	YN84_MYCTU	P96889	mycobacteri
23	66.5	9.6	225	1	KDPE_ECOLI	P21866	escherichia
24	66.5	9.6	1370	1	Z261_HUMAN	Q14202	homo sapien
25	66	9.5	427	1	SYH_MYCLE	P46696	mycobacteri
26	66	9.5	549	1	GYRA_MYCKA	Q49608	mycobacteri
27	66	9.5	987	1	EPB4_MOUSE	P54761	mus musculu
28	66	9.5	1407	1	RPOC_BUCAL	P57145	buchera ap
29	66	9.5	2204	1	RRPL_NDVB	P11205	newcastle d
30	65.5	9.5	227	1	UNG_MYCTU	P95119	mycobacteri
31	65.5	9.5	285	1	YW40_PSEAE	Q9H200	pseudomonas
32	65.5	9.5	706	1	GYS_NEUCR	O93869	neurospora
33	65.5	9.5	761	1	AFG3_YEAST	P39925	saccharomyc

```

RESULT.  2
GUNC_CELFI
ID      GUNC_CELFI
AC      P14090;
STANDARD;
PRT;    1101 AA.

```



RP CHARACTERIZATION.  
RX MEDLINE=96413282; PubMed=8816444;  
RA Methot N., Song M.S., Sonenberg N.;  
RT "A region rich in aspartic acid, arginine, tyrosine, and glycine  
RT (DRYG) mediates eukaryotic initiation factor 4B (eIF4B) self-  
RT association and interaction with eIF3.";  
RL MOL. Cell. Biol. 16:5328-5334(1996).  
CC -!- FUNCTION: REQUIRED FOR THE BINDING OF MRNA TO RIBOSOMES. FUNCTIONS  
CC IN CLOSE ASSOCIATION WITH EIF4-F AND EIF4-A. BINDS NEAR THE 5'-  
CC TERMINAL CAP OF MRNA IN PRESENCE OF EIF4-F AND ATP. PROMOTES THE  
CC ATPASE ACTIVITY AND THE ATP-DEPENDENT RNA UNWINDING ACTIVITY OF  
CC BOTH EIF4-A AND EIF4-F.  
CC -!- SUBUNIT: SELF-ASSOCIATES AND INTERACTS WITH EIF3 P170 SUBUNIT.  
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X55733; CAA39265.1; -  
CC PIR; S12566; S12566.  
CC MIM; 603928; -  
CC InterPro: IPR000504; RRM.  
CC Pfam: PF00076; rrm; 1.  
CC SMART; SM00360; RRM; 1.  
CC PROSITE; PS0102; RRM; 1.  
CC PROSITE; PS00030; RRM\_RNP\_1; 1.  
CC Protein biosynthesis; Initiation factor; RNA-binding.  
CC DOMAIN 96 173 RNA-BINDING (RRM).  
CC DOMAIN 164 356 DRYG.  
CC SEQUENCE 611 AA; 69224 MW; 5EAD0891694D00D9 CRC64;  
CC  
Query Match 10.8%; Score 75; DB 1; Length 611;  
Best Local Similarity 25.08; Pred. No. 7;  
Matches 39; Conservative 15; Mismatches 38; Indels 64; Gaps 9;  
QY 14 PVELRYETRDYAVRMTFLPCDAPVTWAFGRELLEDGINRPSGDGVH-----IAP 65  
DB 37 PVSWAEDDD-----LEGDVSTW-----HSNDDVYRAPIDRSILP 74  
QY 66 TPCEGLSDVIRLQVGADRALFRAGAPLVAF-----DRTDKSV----- 105  
DB 75 TAPRAAREPNI-----DRSRL-PKSPPTATPLGNLPYDVTESIKFFRGLNLSAVRLP 127  
QY 106 --PLCQEQTLG----DFEDSLAALGKILAEQNAQ 135  
DB 128 REPSNPERLKGFGYAEFED-LDSLLSALSNEESLG 162  
RESULT 4  
CAH3\_RAT STANDARD; PRT; 259 AA.  
ID CAH3\_RAT  
AC P14141; O54961;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Carbonic anhydrase III (EC 4.2.1.1) (Carbonate dehydratase III) (CA-  
DE III).  
GN CAS.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89166882; PubMed=2852973;  
RA Kelly C.D., Carter N.D., Jeffery S., Edwards Y.H.;  
RT "Characterisation of cDNA clones for rat muscle carbonic anhydrase

RT III.";  
RL Biosci. Rep. 8:401-406(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Mallis R.J., Jung C.-H., Chatterjee T.K., Fisher R.A., Thomas J.A.;  
RT "cDNA sequence of rat liver carbonic anhydrase III.";  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
CC [3]  
CC SEQUENCE FROM N.A.  
CC STRAIN=Wistar; TISSUE=Liver;  
CC MEDLINE=R-20361776; PubMed=10900145;  
CC Ikeda M., Ishii Y., Kato H., Akazawa D., Hatsumura M., Ishida T.,  
CC Matsusue K., Yamada H., Oguri K.;  
CC "Suppression of carbonic anhydrase III in rat liver by a  
CC dioxin-related toxic compound, coplanar polychlorinated biphenyl,  
CC 3,3',4,4',5-pentachlorobiphenyl.";  
CC Arch. Biochem. Biophys. 380:159-164(2000).  
CC -!- FUNCTION: REVERSIBLE HYDRATION OF CARBON DIOXIDE.  
CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M22413; AAA0846.1; -  
CC EMBL; AF037072; BAB92558.1; -  
CC EMBL; AF030829; BAB08111.1; -  
CC HSSP; P00918; ICIIM.  
CC InterPro: IPR001148; Carb.anhydrase.  
CC Pfam: PF00194; carb.anhydrase; 1.  
CC ProDom: PD000865; carb.anhydrase; 1.  
CC PROSITE; PS00162; EUK\_CO2\_ANNHYDRASE; 1.  
CC Lyase; Zinc.  
CC INIT\_MET 0 0  
CC METAL 93 93 ZINC (CATALYTIC).  
CC METAL 95 95 ZINC (CATALYTIC).  
CC METAL 118 118 ZINC (CATALYTIC).  
CC CONFLICT 129 130 A -> G (IN REF. 1).  
CC CONFLICT 129 130 FG -> SE (IN REF. 1).  
CC CONFLICT 223 224 KL -> NV (IN REF. 1).  
CC SEQUENCE 259 AA; 29300 MW; 7BF7DA65E9493F21 CRC64;  
CC  
Query Match 10.8%; Score 74.5; DB 1; Length 259;  
Best Local Similarity 26.6%; Pred. No. 2.8;  
Matches 29; Conservative 14; Mismatches 39; Indels 27; Gaps 5;  
QY 24 PYAVRMTFLPCDAPVTWA----FGRELLLDGINRPSGDGVHAP-----TDPEG 70  
DB 86 PYRLRQ-FHL-----HWGSSDDHGHSEHTVDGVKVAELHLVHWPKNYTFGEALKQPDG 138  
QY 71 LSDVSIRLQVGADRALFRAGAPLVAFDRTDKSVPLQGEOTLGDGFS 119  
DB 139 IAVVGIFLKIIGKEGFE-----ILLDALDKIKTKKEAPFNHFDPS 180  
RESULT 5  
CAH3\_MOUSE  
ID CAH3\_MOUSE STANDARD; PRT; 259 AA.  
AC P16015;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Carbonic anhydrase III (EC 4.2.1.1) (Carbonate dehydratase III) (CA-  
DE III).

```

Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Badoock K.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Gordon S.K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies K., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Stalston J.E., Taylor K., Whitehead S., Barrell B.G.;
"Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
Nature 393:537-544(1998).
[2]
SEQUENCE FROM N.A.
STRAIN=CDC 1551 / Oshkosh;
Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
Kolonyak J.F., Nelson W.C., Unaym L.A., Emolaeva M.D., Salzberg S.L.,
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bisbal W.;
"Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: BELONGS TO THE UPF0027 (RTCB) FAMILY.
CC -! CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
FRAMESHIFT IN POSITION 257.
-----CC-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----CC-----
EMBL; Z80225; CAB02365.1; ALT_FRAME.
DR EMBL; AS007103; AAK47022.1; -.
DR TIGR; MT2707; -.
DR TubercuList; RV2631; -.
DR InterPro; IPR001233; UPF0027.
DR Pfam; PF01139; UPF0027; 1.
DR PROSITE; PS01288; UPF0027; 1.
DR KW Hypothetical protein; Complete proteome.
FT CONFLICT 393 393 D -> G (IN REF. 1).
SQ SEQUENCE 432 AA; 45586 MW; 2F1I75CAID9BEED4 CRC64;
Query Match 10.5%; Score 73; DB 1; Length 432;
Best Local Similarity 28.9%; Pred. No. 7.3;
Matches 28; Conservative 14; Mismatches 25; Indels 30; Gaps 6;
QY 24 PYAVRMTHFLPGDAPVFWAFG-----RELLDGNRPSCGDGVHTAPTDEGLSDVS 75
Db 10 PGIVRSAYNP---DVHWGYGFPTGGVAANTVDNDGVVSPGVG-----FDIS 54
QY 76 --IRLQV--GADRALFRAGAPPLVAFLDRTKSVPLG 108
Db 55 CGVRLLVGEGLDREELQ---PRLPAVMMDRLDRAIPRG 88
RESULT 7
YAGR_ECOLI ID YAGR_ECOLI STANDARD; PRT; 732 AA.
AC P77489;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yagr.
GN YAGR OR B0284.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
```



```

RL  Proc. Natl. Acad. Sci. U.S.A. 83:9571-9575(1986).
RN  [3]
RP  PARTIAL SEQUENCE FROM N.A.
RX  MEDLINE=88056301; PubMed=2824285;
RA  Lloyd J., Brownson C., Tweedie S., Charlton J., Edwards V.H.;
RT  "Human muscle carbonic anhydrase: gene structure and DNA methylation
RT  patterns in fetal and adult tissues.";
RL  Genes Dev. 1:594-602(1987).
CC  -!- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.
CC  -!- CATALYTIC ACTIVITY: H(2)CO(3) -> CO(2) + H(2)O.
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC  -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC  -!- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE
CC  FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M29458; AA52293.1; .
DR  EMBL; M29452; .; NOT_ANNOTATED_CDS.
DR  PIR; A25850; A25850.
DR  PIR; A26658; A26658.
DR  PIR; A26690; A26690.
DR  HSSP; P00918; 1CIM.
DR  USC-2DPAGE; P07451; HUMAN.
DR  MIM; 114750; .
DR  InterPro; IPR001148; Carb_anhydrase.
DR  Pfam; PF00194; carb_anhydrase; 1.
DR  ProDom; PD000865; Carb_anhydrase; 1.
DR  PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1.
KW  Lyase; Zinc.
FT  INIT_MET 0 0 BY SIMILARITY.
FT  METAL 93 93 ZINC (CATALYTIC).
FT  METAL 95 95 ZINC (CATALYTIC).
FT  METAL 118 118 ZINC (CATALYTIC).
FT  CONFLICT 30 30 I -> V (IN REF. 2).
SQ  SEQUENCE 259 AA; 29440 MW; EFBC56CF7F721E4E CRC64;

Query Match 10.2%; Score 70.5; DB 1; Length 259;
Best Local Similarity 24.1%; Pred. No. 7;
Matches 27; Conservative 18; Mismatches 34; Indels 33; Gaps 6;

QY 24 PYAVRMTFHLPGDAPVTWA-----FGRELLLDGINRPSGQGVHIAPTDP----- 68
Db 86 PYRLRQ-FHL-----HWGSSDDRGSEHTVDGVKYAA---ELHLVHNPKYNTFKKALKQ 135
QY 69 -EGLSDVSIQLQVADRALFRACAPLVAFLDRTKSVPLGQQTGLGDPDS 119
Db 136 RDGIAVIGIFLKIGHENGFEQ-----IFLDALDKIKTKGREAPTFTKFDPS 180

RESULT 10
LMA5_MOUSE
ID LMA5_MOUSE STANDARD; PRT; 3718 AA.
AC Q61001; Q9JHQ6;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-92 FROM N.A.
RA Timpl R., Sasaki T.;

```

```

RT  Completion of the N-terminal sequence of the murine Laminin alpha 5
RT  chain.;
RL  Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE OF 84-3718 FROM N.A.
RX  STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX  MEDLINE=96081906; PubMed=7499364;
RA  Miner J.H., Lewis R.M., Sanes J.R.;
RT  "Molecular cloning of a novel laminin chain, alpha 5, and widespread
RT  expression in adult mouse tissues.";
RL  J. Biol. Chem. 270:28523-28526(1995).
RN  [3]
RP  REVISIONS.
RA  Miner J.H., Lewis R.M., Sanes J.R.;
RA  Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC  -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC  IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC  CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC  WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC  -!- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
CC  ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC  -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC  DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC  TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC  COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC  -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC  MEMBRANES (MAJOR COMPONENT).
CC  -!- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
CC  KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
CC  AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC  -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC  WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC  -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC  -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC  -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC  -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC  -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AJ293593; CAB99255.1; .
DR  EMBL; U37501; AAC53430.1; .
DR  HSSP; P02468; LTLE.
DR  MGD; MGI:105382; Lama5.
DR  InterPro; IPR000561; EGF-like.
DR  InterPro; IPR001886; Lamin.
DR  InterPro; IPR000034; Laminin_B.
DR  InterPro; IPR002049; Laminin_EGF.
DR  InterPro; IPR001791; Laminin_G.
DR  Pfam; PF00052; laminin_B; 1.
DR  Pfam; PF00053; laminin_EGF; 19.
DR  Pfam; PF00054; laminin_G; 2.
DR  Pfam; PF00055; laminin_Nterm; 1.
DR  PRINTS; P000011; EGFLAMININ.
DR  ProDom; PD002082; LamNT; 1.
DR  ProDom; PD003031; Laminin_B; 1.
DR  SMART; SM00180; EGF_Lam; 17.
DR  SMART; SM00001; EGF_Like; 2.
DR  SMART; SM00281; LamB; 1.
DR  SMART; SM00282; LamG; 5.
DR  SMART; SM00136; LamNT; 1.
DR  PROSITE; PS00022; EGF_1; 19.
DR  PROSITE; PS01186; EGF_2; 3.
DR  PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
DR  PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW  Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW  Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

```







RA Wagner M.J., Packer J.C.L., Howe C.J., Bendall D.S.;  
 RT "Some characteristics of cytochrome f in the cyanobacterium Phormidium  
 RT lamiosum: its sequence and charge properties in the reaction with  
 RT plastocyanin.";  
 RL Biochim. Biophys. Acta 1276:246-252(1996).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 46-295.  
 RX MEDLINE=99352179; PubMed=10423236;  
 RA Carrell C.J., Schlarb B.G., Bendall D.S., Howe C.J., Cramer W.A.,  
 RA Smith J.L.;  
 RT "Structure of the soluble domain of cytochrome f from the  
 RT cyanobacterium Phormidium lamiosum.";  
 RL Biochemistry 38:9590-9599(1999).  
 CC -!- FUNCTION: TRANSLOCATES PROTONS ACROSS THE THYLAKOID MEMBRANE AND  
 CC TRANSFERS ELECTRONS FROM PHOTOSYSTEM II TO PHOTOSYSTEM I. IT  
 CC RECEIVES ELECTRONS FROM THE RIESKE IRON-SULFUR PROTEIN AND PASSES  
 CC THEM TO PLASTOCYANIN; THIS FUNCTION IS VERY SIMILAR TO THAT OF  
 CC MITOCHONDRIAL CYTOCHROME C1.  
 CC -!- SUBUNIT: MEMBER OF THE CYTOCHROME B6/F COMPLEX INCLUDING  
 CC CYTOCHROME B6, CYTOCHROME F AND PROBABLY AN IRON SULFUR PROTEIN.  
 CC -!- SUBCELLULAR LOCATION: THYLAKOID MEMBRANE (PROBABLE).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; Y09612; CAA70824.1; -;  
 DR PDB; 1C13; 31-MAY-00.  
 DR InterPro; IPR002325; ApocytL\_F.  
 DR InterPro; IPR000345; CytC\_heme\_bind.  
 DR Pfam; PF01333; Apocytochrome.F; 1.  
 DR PRINTS; PR00610; CYTOCHROME.F.  
 DR PROSITE; PS00190; CYTOCHROME.C; 1.  
 KW Electron transport; Heme; Photosynthesis; Photosystem I;  
 KW Photosystem II; Signal; 3D-structure.  
 FT SIGNAL 1 45  
 FT CHAIN 46 338 APOCYTOCHROME F.  
 FT METAL 46 46 IRON (HEME AXIAL LIGAND).  
 FT BINDING 66 66 HEME (COVALENT).  
 FT BINDING 69 69 HEME (COVALENT).  
 FT METAL 70 70 IRON (HEME AXIAL LIGAND).  
 SQ SEQUENCE 338 AA; 36069 MW; BD3D52036181FB50 CRC64;  
 Query Match 10.0%; Score 69.5; DB 1; Length 338;  
 Best Local Similarity 22.3%; Pred. No. 12;  
 Matches 37; Conservative 20; Mismatches 48; Indels 61; Gaps 8;  
 QY 12 KIPVELRYETR-----DPYA-----VRMTFHLPGDA-----PV-----TWAFGRE 46  
 Db 133 RIPEEMKEEVGSVLFQPYADDKQNVLVGLPLPGDYEIVFPVLPSPNPNATKNSVAFCKY 192  
 QY 47 LLLDINPSPGDGVDHIAPTDPEGLSDVSIIRLOVGADRALFRAGAPPLVAFLDRTD---- 102  
 Db 193 SIHLGANRGRG-----QIYPTGEK-----SNNVAVNNSAAGVITAIAKADGSA 236  
 QY 103 -----KSVPLGQEQTLGDFEELSLEAALGKILAEENQAG 135  
 Db 237 EVKIRTEGTTIVDKIPAGPELVISEG-----EVAAGAALTNNPNVG 279  
 RESULT 14  
 ID HXK1\_SCHPO STANDARD; PRT; 484 AA.  
 AC Q09756;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Hexokinase 1 (EC 2.7.1.1).  
 GN HXK1 OR SPAC24H6.04.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96140736; PubMed=8549830;  
 RA Petit T., Blazquez M.A., Gancedo C.;  
 RT "Schizosaccharomyces pombe possesses an unusual and a conventional  
 RT hexokinase: biochemical and molecular characterization of both  
 RT hexokinases.";  
 RL FEBS Lett. 378:185-189(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: HAS LOW AFFINITY FOR GLUCOSE AND SOME OF ITS ANALOGS.  
 CC -!- CATALYTIC ACTIVITY: ATP + D-hexose = ADP + D-hexose 6-phosphate.  
 CC -!- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.  
 CC -!- SUBUNIT: MONOMER.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X92894; CAA63487.1; -;  
 DR EMBL; Z54142; CAA90848.1; -;  
 DR HSSP; P19367; LHKC.  
 DR InterPro; IPR001312; Hexokinase.  
 DR Pfam; PF00349; hexokinase; 1.  
 DR PRINTS; PR00475; HEXOKINASE.  
 DR PRODOM; PD001109; Hexokinase; 1.  
 DR PROSITE; PS00378; HEXOKINASES; 1.  
 KW Transferase; Kinase; Glycolysis; ATP-binding; Multigene family.  
 FT BINDING 115 115 ATP (BY SIMILARITY).  
 FT DOMAIN 154 180 GLUCOSE-BINDING (POTENTIAL).  
 SQ SEQUENCE 484 AA; 53597 MW; 165500F1966BB0F CRC64;  
 Query Match 10.0%; Score 69.5; DB 1; Length 484;  
 Best Local Similarity 26.2%; Pred. No. 19;  
 Matches 33; Conservative 14; Mismatches 52; Indels 27; Gaps 6;  
 QY 8 ELSFKIPVELRYETRDPAVRMTFHL---PGDAPV---TWAFGRELLDGINRPS----- 56  
 Db 35 EQQFTIPTELLHRVTDTRFVSELYKGLTTNPGDVPVMPVTWIICTP---DGNHGSYALDL 91  
 QY 57-GDGDVHIAPTDEGLSDVSI-----RLOVGADRALFRAGAPPLVAFLDRTDKSVPL 107  
 Db 92 GGTNRVCACAVEQVGNGKFDITOSKYRLPOELKVGTRFALFDYIADCIKKFVE-----EVHP 147  
 QY 108 GQEQTL 113  
 Db 148 GKSQL 153  
 RESULT 15  
 ID YT17\_MYCTU STANDARD; PRT; 602 AA.  
 AC Q10966; Q10967;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 65.8 kDa protein Rv2917.  
 GN RV2917 OR MT2985 OR MTCV338.05.

Job time: 946 sec

```
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: STRONG, TO M.LEPRAE ML1624.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z74697; CAA98979.1; ALT_INIT.
CC EMBL; AE007120; AAK47311.1; -.
CC TIGR; MT2985; -.
CC TuberculList; RV2917; -.
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001410; DEAD.
CC SMART; SM00382; AAA; 1.
CC SMART; SM00487; DEXdc; 1.
CC Hypothetical protein; Complete proteome.
KW CONFLICT 352 352 S -> G (IN REF. 2).
FT CONFLICT 570 570 R -> L (IN REF. 2).
SQ SEQUENCE 602 AA; 65824 MW; 321718F78BE746DC CRC64;

Query Match 10.0%; Score 69.5; DB 1; Length 602;
Best Local Similarity 30.5%; Pred. No. 24;
Matches 40; Conservative 11; Mismatches 57; Indels 23; Gaps 7;

QY 17 LRYETRDYAVRMTFHLPGDAPVTW-----AFG---RELLLDGINRPSGGDGVHIAPTD 67
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 148 VRTEARKTLVVDFEIHGGDAK-TWGDAIREAFGDATFRRLALTGTPFRSDSPFPVSYQ 206
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 68 PEGLSDVYSIRLQV----GADRALFRAGAPPLVAFDRDKSVPLGQQTLCDFEDSLEAA 123
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 207 PD--ADGVLSRQADHTYGYAEALADGVVRPV-FLAYS-----GQARWRDSAGEYEAR 257
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 124 LGKILAEQNA 134
   ||:|:| | | |
Db 258 LGEPLSAEQTA 268
```

Search completed: July 18, 2002, 14:31:57

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 14:30:57 ; Search time 140.15 Seconds  
(without alignments)  
166.638 Million cell updates/sec

Title: US-09-749-185-9  
Perfect score: 692  
Sequence: 1 MSFLYSEELSKIPVELRYE.....FEDSLAALGKILAEQNAG 135

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	692	100.0	145	2	Q9F9B5	Q9F9B5	streptomyces
2	606	87.6	136	2	P95733	P95733	streptomyces
3	548	79.2	136	2	Q9X9U2	Q9X9U2	streptomyces
4	544	78.6	135	2	Q9F9B7	Q9F9B7	streptomyces
5	446	64.5	135	2	Q9F9B6	Q9F9B6	streptomyces
6	246	35.5	159	2	Q9L268	Q9L268	streptomyces
7	237	34.2	142	2	Q9S2F7	Q9S2F7	streptomyces
8	168	24.3	156	2	Q9FC07	Q9FC07	streptomyces
9	159	23.0	138	2	Q9X7M8	Q9X7M8	streptomyces
10	132.5	19.1	142	2	Q9X7R1	Q9X7R1	streptomyces
11	103.5	15.0	126	2	Q9KRC9	Q9KRC9	streptomyces
12	81	11.7	537	2	Q9ACY2	Q9ACY2	streptomyces
13	81	11.7	1370	11	Q9JLM4	Q9JLM4	mus musculus
14	80	11.6	514	16	Q9JVC9	Q9JVC9	neisseria m
15	79	11.4	341	16	Q9CKD1	Q9CKD1	pasteurella
16	76.5	11.1	313	17	Q9HSS8	Q9HSS8	halobacteri

17	76.5	11.1	347	5	Q9BLC6	Q9BLC6 cellana gra
18	76.5	11.1	354	5	Q9BLC7	Q9BLC7 aplysia kur
19	76	11.0	514	16	Q9K0C4	Q9K0C4 neisseria m
20	75.5	10.9	591	2	Q9RKD0	Q9RKD0 streptomyce
21	75.5	10.9	818	16	Q9AAG7	Q9AAG7 caulobacter
22	75.5	10.9	876	2	Q9EXJ5	Q9EXJ5 escherichia
23	75.5	10.9	902	16	Q9I742	Q9I742 pseudomonas
24	75	10.8	792	2	Q9L2I8	Q9L2I8 streptomyce
25	75	10.8	2365	2	Q93HH9	Q93HH9 streptomyce
26	74.5	10.8	291	16	P71922	P71922 mycobacteri
27	74.5	10.8	1348	16	Q988Y5	Q988Y5 rhizobium l
28	74.5	10.8	4280	5	Q9UB29	Q9UB29 caenorhabdi
29	74.5	10.8	4450	5	Q9UB28	Q9UB28 caenorhabdi
30	74	10.7	347	16	Q9A6R1	Q9A6R1 caulobacter
31	74	10.7	402	2	Q93JF9	Q93JF9 streptomyce
32	74	10.7	456	10	Q9SR03	Q9SR03 arabidopsis
33	73.5	10.6	346	16	Q34788	Q34788 bacillus su
34	73.5	10.6	934	2	Q9K452	Q9K452 streptomyce
35	73	10.5	287	11	Q99KB7	Q99KB7 mus musculu
36	73	10.5	945	16	Q98IM7	Q98IM7 rhizobium l
37	73	10.5	1234	12	Q99FJ0	Q99FJ0 porcine tes
38	73	10.5	1337	16	P76465	P76465 escherichia
39	73	10.5	2207	12	Q99FI9	Q99FI9 porcine tes
40	72.5	10.5	260	11	Q9ERN8	Q9ERN8 mus musculu
41	72.5	10.5	343	7	Q31271	Q31271 rattus norv
42	72.5	10.5	348	7	Q77949	Q77949 rattus norv
43	72.5	10.5	462	2	Q93ED0	Q93ED0 rhizobium l
44	72.5	10.5	657	16	Q9RWE6	Q9RWE6 deinococcus
45	72	10.4	967	12	Q99FI5	Q99FI5 porcine tes

ALIGNMENTS

RESULT: 1

Q9F9B5 PRELIMINARY: PRT; 145 AA.  
ID Q9F9B5;  
AC Q9F9B5;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DE 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
DE SSGA.  
GN SSGA.  
OS Streptovorticillum netropsis (Streptovorticillum flavopersicus).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=55404;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA van Wezel G.P., Rousseau C., Kraal B.;  
RT "Cloning and sequencing of the Streptomycetes netropsis ssgA gene";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF195772; AAG28483.1; -;  
SQ SEQUENCE 145 AA; 15838 MW; 92AC0F30DF3D2620 CRC64;

Query Match 100.0%; Score 692; DB 2; Length 145;  
Best Local Similarity 100.0%; Pred. No. 8.3e-62;  
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MSFLYSEELSKIPVELRYEYTRDPYAVRMTFHLPGDAPVTWAFGRELIDGINRPSGDCG 60  
Db 11 MSFLYSEELSKIPVELRYEYTRDPYAVRMTFHLPGDAPVTWAFGRELIDGINRPSGDCG 70  
Qy 61 VHIAPTDEGLSDVSRVQVQADRALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDEL 120  
Db 71 VHIAPTDEGLSDVSRVQVQADRALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDEL 130  
Qy 121 EALGKILAEQNAG 135  
Db 131 EALGKILAEQNAG 145



||||| ||| :| |||||:|:| | ||||| ||||| |||||:| ||| |  
Db 61 VHIAADPETFGVLRQVSDQAMFRVGTAPLVAFLDRTDKIVPLGQERSLADFALL 120  
QY 121 EAALGKILAEQNAQ 135  
Db 121 DEALDRILAEQNAQ 135  
RESULT 5  
Q9F9B6 PRELIMINARY: PRT: 135 AA.  
AC Q9F9B6  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE SSGA.  
DE SSGA.  
OS Streptomyces albus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA van Wezel G.P., Rousseau C., Kraal B.;  
RT "Cloning and sequencing of the Streptomyces albus ssgA gene."  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF195771; AAG28482.1; -  
SQ SEQUENCE 135 AA; 14735 MW; 0FCBF4BDB2BA201B CRC64;

Query Match 64.5%; Score 446; DB 2; Length 135;  
Best Local Similarity 63.0%; Pred. No. 3.6e-37;  
Matches 85; Conservative 20; Mismatches 30; Indels 0; Gaps 0;  
QY 1 MSFLVSEELSFKIPVELRYETRDPAVAVMTFHLPGDAPVTWAFGRELLEDGINRPSGDGD 60  
Db 1 MSFLVSEELAFRIPVELRYETVDPYAVRLTFHLPGDAPVTWVFGRELLVEGVLDAGDGD 60  
QY 61 VHIAPTDPEGLSDVSIRLQVGADRALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDSL 120  
Db 61 VRVCPVQTATREVIHTLVQVSEQALFRVGKAPLVAFLDRTDQGLSGRAHADFDSHL 120

QY 121 EAALGKILAEQNAQ 135  
Db 121 DDALNRSLAEQNAQ 135  
RESULT 6  
Q9L268 PRELIMINARY: PRT: 159 AA.  
AC Q9L268  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PUTATIVE REGULATOR.  
GN SCL2.31.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);

RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL: AL137778; CAB70943.1; -  
SQ SEQUENCE 159 AA; 17472 MW; BA41013F940D7315 CRC64;  
Query Match 35.5%; Score 246; DB 2; Length 159;  
Best Local Similarity 43.1%; Pred. No. 5.1e-17;  
Matches 56; Conservative 15; Mismatches 55; Indels 4; Gaps 1;  
QY 1 MSFLVSEELSFKIPVELRYETRDPAVAVMTFHLPGDAPVTWAFGRELLEDGINRPSGDGD 60  
Db 33 LRLVSSSESLPAGRLYDTADPYAVHATFTTGAEETVEWVFARDLLEGLHRPTGTDG 92  
QY 61 VHIAPTDPEGLSDVSIRLQVGADRALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDSL 120  
Db 93 VRVWFSRSHGQGVVCIALLSPEGEALLEAPARALEFLKRTDAAVPPGTEHRRFD---L 148  
QY 121 EAALGKILAE 130  
Db 149 DQELSHILAE 158

RESULT 7  
Q9S2F7 PRELIMINARY: PRT: 142 AA.  
AC Q9S2F7  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE PUTATIVE REGULATOR.  
GN SCE19A.24.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K., Harris D.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL: AL096852; CAB51005.1; -  
SQ SEQUENCE 142 AA; 15364 MW; 857862390AA51CCB CRC64;

Query Match 34.2%; Score 237; DB 2; Length 142;  
Best Local Similarity 39.7%; Pred. No. 3.5e-16;  
Matches 50; Conservative 20; Mismatches 56; Indels 0; Gaps 0;  
QY 1 MSFLVSEELSFKIPVELRYETRDPAVAVMTFHLPGDAPVTWAFGRELLEDGINRPSGDGD 60  
Db 17 LRLVSPESGIPVPAFLGYHTDDPYAVHITFTHDSGHPVHTVFAFDLLVEGVFRPSGHGD 76  
QY 61 VHIAPTDPEGLSDVSIRLQVGADRALFRAGAPPLVAFLDRTDKSVPLGQEQTLGDFEDSL 120  
Db 61 VRVWFSRSHGQGVVCIALLSPEGEALLEAPARALEFLKRTDAAVPPGTEHRRFD---L 148

```
Db 77 VRWPKTEGRSVVLVALSSPDGDALLEAPTQVSAMLERLRAVPPGTEGAQLGIDGDL 136
Qy 121 EAALGK 126
   | :
Db 137 AELLAR 142

RESULT 8
Q9FC07 PRELIMINARY; PRT; 156 AA.
AC Q9FC07;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PUTATIVE REGULATOR.
GN SC8A11.03
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapaitte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL391041; CAC01575.1;
SQ SEQUENCE 156 AA; 17188 MW; 7181972EE7F33FB0 CRC64;

Query Match 24.3%; Score 168; DB 2; Length 156;
Best Local Similarity 38.1%; Pred. No. 3.3e-09;
Matches 48; Conservative 15; Mismatches 47; Indels 16; Gaps 5;

Qy 13 IPVELRYETRDYAVRMTFHLPGDAPVTWAFGRRELLDGINRPSGDGDVHIAP---APTDP 69
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 37 VPAEFYDPDHPHLLVIRFAPEGAPPVTHVGRDLLHEGLRTTSGLDGVQVWADPTDRE 96
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 70 GLSDVSIKLYGA--DRALFRAGAPPLVAFLDRTDKSVPLGQEQT-LGDFEDSLEAALGK 126
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 -----TAWLQVNAHGDIATFSLPVELEWIDRTYLVHPAGTESRLG-----TDAFLSK 146
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 127 ILAEEQ 132
   | :
Db 147 LFDPE 152

RESULT 9
Q9X7M8 PRELIMINARY; PRT; 138 AA.
AC Q9X7M8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE REGULATOR.
GN SC5F2A.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaitte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL049387; CAB40672.1;
SQ SEQUENCE 138 AA; 15261 MW; F531BA295144ABA5 CRC64;

Query Match 23.0%; Score 159; DB 2; Length 138;
Best Local Similarity 33.1%; Pred. No. 2.2e-08;
Matches 40; Conservative 13; Mismatches 42; Indels 26; Gaps 3;

Qy 13 IPVELRYETRDYAVRMTFHLPGDAPVTWAFGRRELLDGINRPSGDGDVHIAP--- 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 23 IPATLHYDRADPPFAVRMTFPAPATLEGVEVCWTSRELLIAGMQEPNGHGVRVRPAYD 82
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 66 -----TDPEGLSDVSIKLYGA--DRALFRAGAPPLVAFLDRTDKSVPLGQEQT-LGDFEDS 119
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83 RTVLEFNAPECTAVIHVR-----SGELRFLQAGELVPVGLHQLDLQDLHD 129
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 120 L 120
   |
Db 130 L 130

RESULT 10
Q9X7R1 PRELIMINARY; PRT; 142 AA.
AC Q9X7R1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE REGULATOR.
GN SC5H1.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaitte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL049863; CAB42928.1;
SQ SEQUENCE 138 AA; 15261 MW; F531BA295144ABA5 CRC64;
```



[illegible]

[illegible]

Search completed: July 18, 2002, 14:30:58  
Job time: 947 sec